PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07K 14/435 — A1

(11) International Publication Number: WO 00/59936

(43) International Publication Date: 12 October 2000 (12.10.00)

(21) International Application Number: PCT/US00/08436

(22) International Filing Date: 28 March 2000 (28.03.00)

(30) Priority Data:
60/128,180 6 April 1999 (06.04.99) US

(71) Applicant *(for all designated States except US)*: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 12th floor, 1111 Franklin Street, Oakland, CA 94607-5200 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: GERMAN, Michael, S. [US/US]; 1543 33rd Avenue, San Francisco, CA 94122 (US). LIN, Joseph [US/US]; 3436 Noriega, San Francisco, CA 94122 (US).

(74) Agent: FRANCIS, Carol, L.; Bozicevic, Field & Francis LLP, Suite 200, 200 Middlefield Road, Menlo Park, CA 94025 (US).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: HUMAN NEUROGENIN 3-ENCODING NUCLEOTIDE SEQUENCES

(57) Abstract

The present invention features a human neurogenin3 (Ngn3) polypeptide and nucleotide sequences encoding Ngn3 polypeptides. In a particular aspect, the polynucleotide is the nucleotide sequence of SEQ ID NO:1. In addition, the invention features isolated nucleic acid sequence comprising an Ngn3 promoter, as well as a polynucleotide sequences that hybridize under stringent conditions to SEQ ID NO:1. In related aspects the invention features expression vectors and host cells comprising polynucleotides that encode a human Ngn3 polypeptide. The present invention also relates to antibodies that bind specifically to a human Ngn3 polypeptide, methods for producing human Ngn3 polypeptides, methods for identifying β-cell precursor cells expressing Ngn3, methods for using the Ngn3 gene and the Ngn3 polypeptide to alter cellular differentiation in culture or *in vivo* to produce new β-cells to treat patients with diabetes mellitus, and identification of individuals at risk for diabetes by detecting alteration in Ngn3 coding and regulatory sequences and Ngn3 expression levels.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

HUMAN NEUROGENIN 3-ENCODING NUCLEOTIDE SEQUENCES

FIELD OF THE INVENTION

The invention relates generally to the field of nucleotide sequences encoding transcription factors involved in growth and differentiation, particularly development of pancreatic islet cells.

BACKGROUND OF THE INVENTION

Diabetes mellitus is the third leading cause of death in the U.S. and the leading cause of blindness, renal failure, and amputation. Diabetes is also a major cause of premature heart attacks and stroke and accounts for 15% of U.S. health care costs. Approximately 5% of Americans, and as many as 20% of those over the age of 65, have diabetes.

Diabetes results from the failure of the β-cells in the islets of Langerhans in the endocrine pancreas to produce adequate insulin to meet metabolic needs. Diabetes is categorized into two clinical forms: Type 1 diabetes (or insulin-dependent diabetes) and Type 2 diabetes (or non-insulin-dependent diabetes). Type 1 diabetes is caused by the loss of the insulin-producing β-cells. Type 2 diabetes is a more strongly genetic disease than Type 1 (Zonana & Rimoin, 1976 N. Engl. J. Med. 295:603), usually has its onset alter in life, and accounts for approximately 90% of diabetes in the U.S. Affected individuals usually have both a decrease in the capacity of the pancreas to produce insulin and a defect in the ability to utilize the insulin (insulin resistance). Obesity causes insulin resistance, and approximately 80% of individuals with Type 2 diabetes are clinically obese (greater than 20% above ideal body weight). Unfortunately, about one-half of the people in the U.S. affected by Type 2 diabetes are unaware that they have the disease. Clinical symptoms associated with Type 2 diabetes may not become obvious until late in the disease, and the early signs are often misdiagnosed, causing a delay in treatment and increased complications. While the role of genetics in the etiology of type 2 diabetes is clear, the precise genes involved are largely unknown.

Insulin is made exclusively by the β-cells in the islets of Langerhans in the pancreas. During development, the islet cells, including the β-cells, develop from an undifferentiated precursor within the growing pancreatic bud. As the bud grows, the undifferentiated cells form into ducts, and it is these cells that function as precursors. Duct cells appear to retain

the capacity to differentiate into islet cells throughout life, and when the pancreas is damaged, new islet cells form from the duct cells.

This developmental process is clinically relevant for several reasons. First, the formation of islet cells and especially β-cells is necessary in order to make insulin and control energy metabolism. If the process of β-cell development is in anyway impaired, it predisposes that individual to the later development of diabetes. Therefore genes involved in this process are candidate genes for neonatal diabetes, maturity onset diabetes of the young (MODY) or type 2 diabetes. The sequence of these genes could be used to identify individuals at risk for the development of diabetes, or to develop new pharmacological agents to prevent and treat diabetes.

Second, as discussed above, insulin production is impaired in individuals with diabetes. In type 1 diabetes the impairment is caused by the destruction of the beta-cells, while in type 2 diabetes, insulin production is intact, but inadequate. Treatment of type 1 diabetes, as well as many cases of type 2 diabetes, may involve replacement of the β-cells. While replacement of β-cells may be accomplished in several ways, the development of new β-cells from precursor cells, either in culture or *in vivo* in the patient, would be the most physiologic. To do this, the molecules that control beta-cell differentiation are needed.

For these reasons, the diabetes field has spent considerable effort in attempts to identify islet precursor cells, and to develop methods for differentiating beta-cells in vitro. To date this has been largely unsuccessful. The present invention addresses this problem.

Relevant Literature

A cloned fragment of mouse Ngn3 is described in Sommer et al. 1996 Mol. Cell. Neurosci. 8:221.

cDNA and amino acid sequences of murine Ngn3 and murine mammalian atonal homology 4B (MATH4B) are described at GenBank Accession Nos. U76208 and Y09167, respectively.

cDNA and amino acid sequences of the rat relax transcriptional regulator are described at GenBank Accession No. Y10619.

SUMMARY OF THE INVENTION

The present invention features a human neurogenin3 (Ngn3) polypeptide and nucleotide sequences encoding Ngn3 polypeptides. In a particular aspect, the polynucleotide is the nucleotide sequence of SEQ ID NO:1. In addition, the invention features isolated nucleic acid sequence comprising an Ngn3 promoter, as well as a polynucleotide sequences that hybridize under stringent conditions to SEQ ID NO:1. In related aspects the invention features expression vectors and host cells comprising polynucleotides that encode a human Ngn3 polypeptide. The present invention also relates to antibodies that bind specifically to a human Ngn3 polypeptide, methods for producing human Ngn3 polypeptides, methods for identifying β-cell precursor cells expressing Ngn3, methods for using the Ngn3 gene and the Ngn3 polypeptide to alter cellular differentiation in culture or *in vivo* to produce new β-cells to treat patients with diabetes mellitus, and identification of individuals at risk for diabetes by detecting alteration in Ngn3 coding and regulatory sequences and Ngn3 expression levels.

A primary object of the invention is to provide an isolated human Ngn3 polypeptide-encoding polynucleotide for use in expression of human Ngn3 (e.g, in a recombinant host cell) and for use in, for example, identification of human Ngn3 polypeptide binding compounds (especially those compounds that affect human Ngn3 polypeptide-mediated activity, which compounds can be used to modulate Ngn3 activity).

Another object of the invention is to provide an isolated human Ngn3 polypeptide-encoding polynucleotide for use in generation of non-human transgenic animal models for Ngn3 gene function, particularly "knock-in" Ngn3 non-human transgenic animals characterized by excess or ectopic expression of the Ngn3 gene.

These and other objects, advantages and features of the present invention will become apparent to those persons skilled in the art upon reading the details of the invention more fully set forth below.

The invention will now be described in further detail.

DETAILED DESCRIPTION OF THE INVENTION

Before the present nucleotide and polypeptide sequences are described, it is to be understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors and reagents described as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only,

and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells and reference to "the antibody" includes reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs. Although any methods, devices and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now described.

All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing, for example, the cell lines, vectors, and methodologies which are described in the publications which might be used in connection with the presently described invention. The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

Definitions

"Polynucleotide" as used herein refers to an oligonucleotide, nucleotide, and fragments or portions thereof, as well as to peptide nucleic acids (PNA), fragments, portions or antisense molecules thereof, and to DNA or RNA of genomic or synthetic origin which can be single- or double-stranded, and represent the sense or antisense strand. Where "polynucleotide" is used to refer to a specific polynucleotide sequence (*e.g.* a Ngn3 polypeptide-encoding polynucleotide), "polynucleotide" is meant to encompass polynucleotides that encode a polypeptide that is functionally equivalent to the recited polypeptide, *e.g.*, polynucleotides that are degenerate variants, or polynucleotides that encode biologically active variants or fragments of the recited polypeptide, including polynucleotides having substantial sequence similarity or sequence identity relative to the sequences provided herein. Similarly, "polypeptide" as used herein refers to an oligopeptide, peptide, or protein. Where "polypeptide" is recited herein to refer to an amino acid sequence of a naturally-

occurring protein molecule, "polypeptide" and like terms are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule, but instead is meant to also encompass biologically active variants or fragments, including polypeptides having substantial sequence similarity or sequence identify relative to the amino acid sequences provided herein.

By "antisense polynucleotide" is mean a polynucleotide having a nucleotide sequence complementary to a given polynucleotide sequence (e.g, a polynucleotide sequence encoding an Ngn3 polypeptide) including polynucleotide sequences associated with the transcription or translation of the given polynucleotide sequence (*e.g.*, a promoter of a polynucleotide encoding an Ngn3 polypeptide), where the antisense polynucleotide is capable of hybridizing to an Ngn3 polypeptide-encoding polynucleotide sequence. Of particular interest are antisense polynucleotides capable of inhibiting transcription and/or translation of an Ngn3-encoding polynucleotide either in vitro or in vivo.

"Peptide nucleic acid" as used herein refers to a molecule which comprises an oligomer to which an amino acid residue, such as lysine, and an amino group have been added. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary (template) strand of nucleic acid (Nielsen et al 1993 Anticancer Drug Des 8:53-63).

As used herein, "Ngn3 polypeptide" refers to an amino acid sequence of a recombinant or nonrecombinant polypeptide having an amino acid sequence of i) a native Ngn3 polypeptide, ii) a biologically active fragment of an Ngn3 polypeptide, *iii)* biologically active polypeptide analogs of an Ngn3 polypeptide, or iv) a biologically active variant of an Ngn3 polypeptide. Ngn3 polypeptides of the invention can be obtained from any species, *e.g.*, mammalian or non-mammalian (*e.g.*, reptiles, amphibians, avian (*e.g.*, chicken)), particularly mammalian, including human, rodenti (e.g., murine or rat), bovine, ovine, porcine, murine, or equine, preferably rat or human, from any source whether natural, synthetic, semi-synthetic or recombinant. "Human Ngn3 polypeptide" refers to the amino acid sequences of isolated human Ngn3 polypeptide obtained from a human, and is meant to include all naturally-occurring allelic variants, and is not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

As used herein, "antigenic amino acid sequence" means an amino acid sequence that, either alone or in association with a carrier molecule, can elicit an antibody response in a mammal.

A "variant" of a human Ngn3 polypeptide is defined as an amino acid sequence that is altered by one or more amino acids. The variant can have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant can have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Similar minor variations can also include amino acid deletions or insertions, or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological or immunological activity can be found using computer programs well known in the art, for example, DNAStar software.

A "deletion" is defined as a change in either amino acid or nucleotide sequence in which one or more amino acid or nucleotide residues, respectively, are absent as compared to an amino acid sequence or nucleotide sequence of a naturally occurring Ngn3 polypeptide.

An "insertion" or "addition" is that change in an amino acid or nucleotide sequence which has resulted in the addition of one or more amino acid or nucleotide residues, respectively, as compared to an amino acid sequence or nucleotide sequence of a naturally occurring Ngn3 polypeptide.

A "substitution" results from the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively as compared to an amino acid sequence or nucleotide sequence of a naturally occurring Ngn3 polypeptide.

The term "biologically active" refers to human Ngn3 polypeptide having structural, regulatory, or biochemical functions of a naturally occurring Ngn3 polypeptide. Likewise, "immunologically active" defines the capability of the natural, recombinant or synthetic human Ngn3 polypeptide, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "derivative" as used herein refers to the chemical modification of a nucleic acid encoding a human Ngn3 polypeptide or the encoded human Ngn3 polypeptide. Illustrative of such modifications would be replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative would encode a polypeptide which retains essential biological characteristics of a natural Ngn3 polypeptide.

As used herein the term "isolated" is meant to describe a compound of interest (e.g., either a polynucleotide or a polypeptide) that is in an environment different from that in which the compound naturally occurs. "Isolated" is meant to include compounds that are within samples that are substantially enriched for the compound of interest and/or in which the compound of interest is partially or substantially purified.

As used herein, the term "substantially purified" refers to a compound (e.g., either a polynucleotide or a polypeptide) that is removed from its natural environment and is at least 60% free, preferably 75% free, and most preferably 90% free from other components with which it is naturally associated.

"Stringency" typically occurs in a range from about Tm-5°C (5°C below the Tm of the probe) to about 20°C to 25°C below Tm. As will be understood by those of skill in the art, a stringency hybridization can be used to identify or detect identical polynucleotide sequences or to identify or detect similar or related polynucleotide sequences.

The term "hybridization" as used herein shall include "any process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs 1994 Dictionary of Biotechnology, Stockton Press, New York NY). Amplification as carried out in the polymerase chain reaction technologies is described in Dieffenbach et al. 1995, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY.

By "transformation" is meant a permanent or transient genetic change, preferably a permanent genetic change, induced in a cell following incorporation of new DNA (i.e., DNA exogenous to the cell). Genetic change can be accomplished either by incorporation of the new DNA into the genome of the host cell, or by transient or stable maintenance of the new DNA as an episomal element. Where the cell is a mammalian cell, a permanent genetic change is generally achieved by introduction of the DNA into the genome of the cell.

By "construct" is meant a recombinant nucleic acid, generally recombinant DNA, that has been generated for the purpose of the expression of a specific nucleotide sequence(s), or is to be used in the construction of other recombinant nucleotide sequences.

By "operably linked" is meant that a DNA sequence and a regulatory sequence(s) are connected in such a way as to permit gene expression when the appropriate molecules (e.g., transcriptional activator proteins) are bound to the regulatory sequence(s).

By "operatively inserted" is meant that a nucleotide sequence of interest is positioned adjacent a nucleotide sequence that directs transcription and translation of the introduced

nucleotide sequence of interest (*i.e.*, facilitates the production of, e.g., a polypeptide encoded by an Ngn3 sequence).

By "Ngn3 associated disorder" is meant a physiological condition or disease associated with altered Ngn3 function (e.g., due to aberrant Ngn3 expression or a defect in Ngn3 expression or in the Ngn3 protein). Such Ngn3 associated disorders can include, but are not necessarily limited to, disorders associated with reduced levels of insulin or the ability to utilize insulin (e.g., hyperglycemia, diabetes (e.g., Type 1 and Type 2 diabetes, and the like).

By "subject" or "patient" is meant any mammalian subject for whom diagnosis or therapy is desired, particularly humans. Other subjects may include cattle, dogs, cats, guinea pigs, rabbits, rats, mice, horses, and so on. Of particular interest are subjects having an Ngn3-associated disorder that is amenable to treatment (*e.g.*, to mitigate symptoms associated with the disorder) by expression of either Ngn3-encoding nucleic acid in a cell of the subject (*e.g.*, by introduction of the Ngn3-encoding nucleic acid into the subject *in vivo*, or by implanting Ngn3-expressing cells (*e.g.*, β-cell precursors) or nearly developed or mature β-cells cultured from Ngn3-expressing cells into the subject, which cells produce insulin).

The term "transgene" is used herein to describe genetic material which has been or is about to be artificially inserted into the genome of a mammalian, particularly a mammalian cell of a living animal.

By "transgenic organism" is meant a non-human organism (e.g., single-cell organisms (e.g., yeast), mammal, non-mammal (e.g., nematode or Drosophila)) having a non-endogenous (i.e., heterologous) nucleic acid sequence present as an extrachromosomal element in a portion of its cells or stably integrated into its germ line DNA.

By "transgenic animal" is meant a non-human animal, usually a mammal, having a non-endogenous (i.e., heterologous) nucleic acid sequence present as an extrachromosomal element in a portion of its cells or stably integrated into its germ line DNA (i.e., in the genomic sequence of most or all of its cells). Heterologous nucleic acid is introduced into the germ line of such transgenic animals by genetic manipulation of, for example, embryos or embryonic stem cells of the host animal.

A "knock-out" of a target gene means an alteration in the sequence of the gene that results in a decrease of function of the target gene, preferably such that target gene expression is undetectable or insignificant. A knock-out of an Ngn3 gene means that function of the

Ngn3 gene has been substantially decreased so that Ngn3 expression is not detectable or only present at insignificant levels. "Knock-out" transgenics of the invention can be transgenic animals having a heterozygous knock-out of the Ngn3 gene or a homozygous knock-out of the Ngn3 gene. "Knock-outs" also include conditional knock-outs, where alteration of the target gene can occur upon, for example, exposure of the animal to a substance that promotes target gene alteration, introduction of an enzyme that promotes recombination at the target gene site (e.g., Cre in the Cre-lox system), or other method for directing the target gene alteration postnatally.

A "knock-in" of a target gene means an alteration in a host cell genome that results in altered expression (e.g., increased (including ectopic) or decreased expression) of the target gene, e.g., by introduction of an additional copy of the target gene, or by operatively inserting a regulatory sequence that provides for enhanced expression of an endogenous copy of the target gene. "Knock-in" transgenics of the invention can be transgenic animals having a heterozygous knock-in of the Ngn3 gene or a homozygous knock-in of the Ngn3 gene. "Knock-ins" also encompass conditional knock-ins.

Overview of the Invention

The present invention is based upon the identification and isolation of a polynucleotide sequence encoding a human neurogenin3 (Ngn3) polypeptide, as well as the human and murine Ngn3 promoters. Accordingly, the present invention encompasses such human Ngn3 polypeptide-encoding polynucleotides, as well as human Ngn3 polypeptides encoded by such polynucleotides. Expression of Ngn3 is linked to pancreatic development. Specifically, Ngn3 expression is the earliest available marker of cells that will develop into islet cells. Because Ngn3 expression is extinguished before the cells are completely differentiated, Ngn3 uniquely marks precursor cells. The proof that these are islet cell precursors is based on three pieces of evidence: 1) Expression pattern. Ngn3 cells are seen scattered through the pancreatic duct cells, with a smaller number present adjacent to the ducts. 2) Timing. The appearance of the Ngn3-positive cells parallels the formation of new islet cells during development. 3) Ngn3-positive cells co-express other endocrine transcription factors, including the β-cell transcription factor Nkx-6.1. Nkx6.1 is known to be expressed in β-cells and β-cell precursors at this stage of pancreatic development, and the knock-out of the Nkx-6.1 gene in

mice causes a specific defect in β-cell development, but no defect in the formation of other pancreatic cells (see, *e.g.*, WO 99/05258).

The invention also encompasses the use of the polynucleotides disclosed herein to facilitate identification and isolation of polynucleotide and polypeptide sequences having homology to a human Ngn3 polypeptide of the invention. The human Ngn3 polypeptides and polynucleotides of the invention are also useful in the identification of human Ngn3 polypeptide-binding compounds, particularly human Ngn3 polypeptide-binding compounds having human Ngn3 polypeptide agonist or antagonist activity. In addition, the human Ngn3 polypeptides and polynucleotides of the invention are useful in the diagnosis, prevention and treatment of disease associated with human Ngn3 polypeptide biological activity.

The human Ngn3 polypeptide-encoding polynucleotides of the invention can also be used in the development of β-cells in culture and *in vivo*, as a molecular probe with which to determine the structure, location, and expression of the human Ngn3 polypeptide and related polypeptides in mammals (including humans), and to investigate potential associations between disease states or clinical disorders and defects or alterations in human Ngn3 polypeptide structure, expression, or function.

Ngn3 Nucleic Acid

The term "Ngn3 gene" is used generically to designate Ngn3 genes and their alternate forms. "Ngn3 gene" is also intended to mean the open reading frame encoding specific Ngn3 polypeptides, introns, and adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 1 kb beyond the coding region, but possibly further in either direction. The DNA sequences encoding Ngn3 may be cDNA or genomic DNA or a fragment thereof. The gene may be introduced into an appropriate vector for extrachromosomal maintenance or for integration into the host.

The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons (*e.g.*, sequences encoding open reading frames of the encoded polypeptide) and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns removed by nuclear RNA splicing, to create a continuous open reading frame encoding the Ngn3 polypeptide.

While other genomic Ngn3 sequences of other sources may have non-contiguous open reading frames (*e.g.*, where introns interrupt the protein coding regions), the human genomic Ngn3 sequence has no introns interrupting the coding sequence. A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It may further include the 3' and 5' untranslated regions found in the mature mRNA. It may further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, *etc.*, including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' or 3' end of the transcribed region. The genomic DNA may be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence.

The sequence of this 5' region, and further 5' upstream sequences and 3' downstream sequences, may be utilized for promoter elements, including enhancer binding sites, that provide for expression in tissues where Ngn3 is expressed. The sequences of the Ngn3 promoter elements of the invention can be based on the nucleotide sequences of any species (*e.g.*, mammalian or non-mammalian (*e.g.*, reptiles, amphibians, avian (*e.g.*, chicken)), particularly mammalian, including human, rodenti (e.g., murine or rat), bovine, ovine, porcine, murine, or equine, preferably rat or human) and can be isolated or produced from any source whether natural, synthetic, semi-synthetic or recombinant.

The tissue specific expression of Ngn3 is useful for determining the pattern of expression, and for providing promoters that mimic the native pattern of expression. Naturally occurring polymorphisms in the promoter region are useful for determining natural variations in expression, particularly those that may be associated with disease. Alternatively, mutations may be introduced into the promoter region to determine the effect of altering expression in experimentally defined systems. Methods for the identification of specific DNA motifs involved in the binding of transcriptional factors are known in the art, *e.g.* sequence similarity to known binding motifs, gel retardation studies, *etc.* For examples, see Blackwell et al. 1995 Mol Med 1:194-205; Mortlock et al. 1996 Genome Res. 6: 327-33; and Joulin and Richard-Foy (1995) Eur J Biochem 232: 620-626.

In one embodiment, the Ngn3 promoter is used to direct expression of genes to islet cell precursors. As discussed below, Ngn3 is expressed in islet cell precursors during development of β-cells. Thus, the developmentally timed expression directed by the Ngn3

promoter can be exploited to facilitate expression of heterologous genes operably linked to the Ngn3 promoter. Exemplary genes of interest that can be expressed from the Ngn3 promoter include, but are not necessarily limited to, genes encoding growth factors or onocogenes (*e.g.*, to expand and/or immortalize the β-cell progenitor population), marker genes (*e.g.*, for marking the precursor cells for selection and/or tracing), reporter genes (*e.g.*, luciferase, CAT, *etc.*, for, *e.g.*, identifying mechanisms for regulating the Ngn3 promoter and/or to search for bioactive agents (*e.g.*, candidate pharmaceutical agents) that regulate the promoter), and the like.

The regulatory sequences may be used to identify *cis* acting sequences required for transcriptional or translational regulation of Ngn3 expression, especially in different tissues or stages of development, and to identify *cis* acting sequences and *trans* acting factors that regulate or mediate Ngn3 expression. Such transcriptional or translational control regions may be operably linked to an Ngn3 gene or other genes in order to promote expression of wild type or altered Ngn3 or other proteins of interest in cultured cells, or in embryonic, fetal or adult tissues, and for gene therapy. Ngn3 transcriptional or translational control regions can also be used to identify extracellular signal molecules that regulate Ngn3 promoter activity, and thus regulate Ngn3 expression and islet cell formation.

The nucleic acid compositions used in the subject invention may encode all or a part of the Ngn3 polypeptides as appropriate. Fragments may be obtained of the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.* For the most part, DNA fragments will be of at least about ten contiguous nucleotides, usually at least about 15 nt, more usually at least about 18 nt to about 20 nt, more usually at least about 25 nt to about 50 nt. Such small DNA fragments are useful as primers for PCR, hybridization screening, *etc.* Larger DNA fragments, *i.e.* greater than 100 nt are useful for production of the encoded polypeptide. For use in amplification reactions, such as PCR, a pair of primers will be used. The exact composition of the primer sequences is not critical to the invention, but for most applications the primers will hybridize to the subject sequence under stringent conditions, as known in the art. It is preferable to choose a pair of primers that will generate an amplification product of at least about 50 nt, preferably at least about 100 nt. Algorithms for the selection of primer sequences are generally known, and are available in commercial software packages.

Amplification primers hybridize to complementary strands of DNA, and will prime towards each other.

The Ngn3 gene is isolated and obtained in substantial purity, generally as other than an intact mammalian chromosome. Usually, the DNA will be obtained substantially free of other nucleic acid sequences that do not include an Ngn3 sequence or fragment thereof, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", *i.e.* flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

The DNA sequences are used in a variety of ways. They may be used as probes for identifying homologs of Ngn3. Mammalian homologs have substantial sequence similarity to one another, *i.e.* at least 75%, usually at least 90%, more usually at least 95% sequence identity. Sequence similarity and sequence identity are calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, *etc.* A reference sequence will usually be at least about 18 nt long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art, such as BLAST, described in Altschul et al. 1990 J Mol Biol 215:403-10. For the purposes of the present application, percent identity for the polynucleotides of the invention is determined using the BLASTN program with the default settings as described at http://www.ncbi.nlm.nih.gov/ cgi-bin/ BLAST/nph-newblast?Jform=0 with the DUST filter selected. The DUST filter is described at http://www.ncbi.nlm.nih.gov/ BLAST/filtered.html.

Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 6XSSC (0.9 M saline/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1XSSC (0.15 M sodium chloride/0.015 M sodium citrate). Sequence identity may be determined by hybridization under high stringency conditions, for example, at 50°C or higher and 0.1XSSC (15 mM saline/0.15 mM sodium citrate). By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes may be any species, *e.g.* primate species, particularly human; rodents, such as rats and mice, canines, felines, bovines, ovines, equines, yeast, *Drosophila*, *Caenhorabditis*, *etc.*

The Ngn3-encoding DNA may be used to identify expression of the gene in a biological specimen. The manner in which one probes cells for the presence of particular

nucleotide sequences, as genomic DNA or RNA, is well established in the literature and does not require elaboration here. mRNA is isolated from a cell sample. mRNA may be amplified by RT-PCR, using reverse transcriptase to form a complementary DNA strand, followed by polymerase chain reaction amplification using primers specific for the subject DNA sequences. Alternatively, mRNA sample is separated by gel electrophoresis, transferred to a suitable support, *e.g.* nitrocellulose, nylon, *etc.*, and then probed with a fragment of the subject DNA as a probe. Other techniques, such as oligonucleotide ligation assays, *in situ* hybridizations, and hybridization to DNA probes arrayed on a solid chip may also find use. Detection of mRNA hybridizing to an Ngn3 sequence is indicative of Ngn3 gene expression in the sample.

The Ngn3 nucleic acid sequence may be modified for a number of purposes, particularly where they will be used intracellularly, for example, by being joined to a nucleic acid cleaving agent, *e.g.* a chelated metal ion, such as iron or chromium for cleavage of the gene; or the like.

The sequence of the Ngn3 locus, including flanking promoter regions and coding regions, may be mutated in various ways known in the art to generate targeted changes in promoter strength, sequence of the encoded protein, *etc.* The DNA sequence or product of such a mutation will be substantially similar to the sequences provided herein, *i.e.* will differ by at least one nucleotide or amino acid, respectively, and may differ by at least two but not more than about ten nucleotides or amino acids. The sequence changes may be substitutions, insertions or deletions. Deletions may further include larger changes, such as deletions of a domain or exon. Other modifications of interest include epitope tagging, *e.g.* with the FLAG system, HA, *etc.* For studies of subcellular localization, fusion proteins with green fluorescent proteins (GFP) may be used. Such mutated genes may be used to study structure-function relationships of Ngn3 polypeptides with other polypeptides (*e.g.*, Nkx-6.1, which is co-expressed with Ngn3), or to alter properties of the proteins that affect their function or regulation. Such modified Ngn3 sequences can be used to, for example, generate the transgenic animals.

Techniques for in vitro mutagenesis of cloned genes are known. Examples of protocols for scanning mutations may be found in Gustin et al., 1993 Biotechniques 14:22 ; Barany, 1985 Gene 37:111-23; Colicelli et al., 1985 Mol Gen Genet 199:537-9; and Prentki et al., 1984 Gene 29:303-13. Methods for site specific mutagenesis can be found in

Sambrook et al., 1989 Molecular Cloning: A Laboratory Manual, CSH Press, pp. 15.3-15.108; Weiner et al., 1993 Gene 126:35-41; Sayers et al., 1992 Biotechniques 13:592-6; Jones and Winistorfer, 1992 Biotechniques 12:528-30; Barton et al., 1990 Nucleic Acids Res 18:7349-55; Marotti and Tomich, 1989 Gene Anal Tech 6:67-70; and Zhu 1989 Anal Biochem 177:120-4.

Ngn3 Transgenic Animals

The Ngn3-encoding nucleic acids can be used to generate genetically modified non-human animals or site specific gene modifications in cell lines. The term "transgenic" is intended to encompass genetically modified animals having a deletion or other knock-out of Ngn3 gene activity, having an exogenous Ngn3 gene that is stably transmitted in the host cells, "knock-in" having altered Ngn3 gene expression, or having an exogenous Ngn3 promoter operably linked to a reporter gene. Of particular interest are homozygous and heterozygous knock-outs of Ngn3.

Transgenic animals may be made through homologous recombination, where the Ngn3 locus is altered. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like. Of interest are transgenic mammals, preferably a mammal from a genus selected from the group consisting of Mus (e.g., mice), Rattus (e.g., rats), Oryctologus (e.g., rabbits) and Mesocricetus (e.g., hamsters). More preferably the animal is a mouse which is defective or contains some other alteration in Ngn3 gene expression or function. Without being held to theory, Ngn3 is a transcription factor that is expressed in islet cell precursors during pancreatic development, transgenic animals having altered Ngn3 gene expression will be useful models of pancreatic development.

A "knock-out" animal is genetically manipulated to substantially reduce, or eliminate endogenous Ngn3 function, preferably such that target gene expression is undetectable or insignificant. Different approaches may be used to achieve the "knock-out". A chromosomal deletion of all or part of the native Ngn3 homolog may be induced. Deletions of the non-coding regions, particularly the promoter region, 3' regulatory sequences, enhancers, or deletions of gene that activate expression of the Ngn3 genes. A functional knock-out may also be achieved by the introduction of an anti-sense construct that blocks expression of the native Ngn3 gene (for example, see Li and Cohen (1996) Cell 85:319-329).

Conditional knock-outs of Ngn3 gene function can also be generated. Conditional knock-outs are transgenic animals that exhibit a defect in Ngn3 gene function upon exposure of the animal to a substance that promotes target gene alteration, introduction of an enzyme that promotes recombination at the target gene site (e.g., Cre in the Cre-loxP system), or other method for directing the target gene alteration.

For example, a transgenic animal having a conditional knock-out of Ngn3 gene function can be produced using the Cre-loxP recombination system (see, e.g., Kilby et al. 1993 Trends Genet 9:413-421). Cre is an enzyme that excises the DNA between two recognition sequences, termed loxP. This system can be used in a variety of ways to create conditional knock-outs of Ngn3. For example, two independent transgenic mice can be produced: one transgenic for an Ngn3. sequence flanked by loxP sites and a second transgenic for Cre. The Cre transgene can be under the control of an inducible or developmentally regulated promoter (Gu et al. 1993 Cell 73:1155-1164; Gu et al. 1994 Science 265:103-106), or under control of a tissue-specific or cell type-specific promoter (e.g., a pancreas-specific promoter or brain tissue-specific promoter). The Ngn3 transgenic is then crossed with the Cre transgenic to produce progeny deficient for the Ngn3 gene only in those cells that expressed Cre during development.

Transgenic animals may be made having an exogenous Ngn3 gene. For example, the transgenic animal may comprise a "knock-in" of an Ngn3 gene, such that the host cell genome contains an alteration that results in altered expression (e.g., increased (including ectopic) or decreased expression) of an Ngn3 gene, e.g., by introduction of an additional copy of the target gene, or by operatively inserting a regulatory sequence that provides for enhanced expression of an endogenous copy of the target gene. "Knock-in" transgenics can be transgenic animals having a heterozygous knock-in of the Ngn3 gene or a homozygous knock-in of the Ngn3. "Knock-ins" also encompass conditional knock-ins.

The exogenous gene introduced into the host cell genome to produce a transgenic animal is usually either from a different species than the animal host, or is otherwise altered in its coding or non-coding sequence. The introduced gene may be a wild-type gene, naturally occurring polymorphism, or a genetically manipulated sequence, for example those previously described with deletions, substitutions or insertions in the coding or non-coding regions. The introduced sequence may encode an Ngn3 polypeptide, or may utilize the Ngn3 promoter operably linked to a reporter gene. Where the introduced gene is a coding sequence, it is

usually operably linked to a promoter, which may be constitutive or inducible, and other regulatory sequences required for expression in the host animal.

Specific constructs of interest include, but are not limited to, anti-sense Ngn3, or a ribozyme based on an Ngn3 sequence, which will block Ngn3 expression, as well as expression of dominant negative Ngn3 mutations, and over-expression of an Ngn3 gene. A detectable marker, such as *lac Z* may be introduced into the Ngn3 locus, where upregulation of expression of the corresponding Ngn gene will result in an easily detected change in phenotype. Constructs utilizing a promoter region of the Ngn3 genes in combination with a reporter gene or with the coding region of Ngn3 are also of interest. Constructs having a sequence encoding a truncated or altered (e.g, mutated) Ngn3 are also of interest.

The modified cells or animals are useful in the study of function and regulation of Ngn3 and other proteins involved the pancreatic β-cell developmental pathway. Such modified cells or animals are also useful in, for example, the study of the function and regulation of genes whose expression is affected by Ngn3, as well as the study of the development of insulin-secreting cells in the pancreas. Thus, the transgenic animals of the invention are useful in identifying downstream targets of Ngn3, as such targets may have a role in the phenotypes associated with defects in Ngn3.

Animals may also be used in functional studies, drug screening, *etc., e.g.* to determine the effect of a candidate drug on islet cell development, on β-cell function and development or on symptoms associated with disease or conditions associated with Ngn3 defects (e.g., on symptoms associated with reduced insulin secretion (*e.g.*, such as that associated with a diabetic syndrome, including Type 2 diabetes). A series of small deletions and/or substitutions may be made in the Ngn3 genes to determine the role of different polypeptide-encoding regions in DNA binding, transcriptional regulation, *etc.* By providing expression of Ngn3 protein in cells in which it is otherwise not normally produced (*e.g.*, ectopic expression), one can induce changes in cell behavior. These animals are also useful for exploring models of inheritance of disorders associated with diabetes, *e.g.* dominant *v.* recessive; relative effects of different alleles and synergistic effects between Ngn3 and other genes elsewhere in the genome.

DNA constructs for homologous recombination will comprise at least a portion of the Ngn3 gene with the desired genetic modification, and will include regions of homology to the target locus. DNA constructs for random integration need not include regions of homology

to mediate recombination. Conveniently, markers for positive and negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art. For various techniques for transfecting mammalian cells, see Keown et al. 1990 Methods in Enzymology 185:527-537.

For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, *e.g.* mouse, rat, guinea pig, *etc.* Such cells are grown on an appropriate fibroblast-feeder layer or grown in the presence of appropriate growth factors, such as leukemia inhibiting factor (LIF). When ES cells have been transformed, they may be used to produce transgenic animals. After transformation, the cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be detected by employing a selective medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination or integration of the construct. Those colonies that are positive may then be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old superovulated females. The ES cells are trypsinized, and the modified cells are injected into the blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant females. Females are then allowed to go to term and the resulting litters screened for mutant cells having the construct. By providing for a different phenotype of the blastocyst and the ES cells, chimeric progeny can be readily detected.

The chimeric animals are screened for the presence of the modified gene. Chimeric animals having the modification (normally chimeric males) are mated with wildtype animals to produce heterozygotes, and the heterozygotes mated to produce homozygotes. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogeneic or congenic grafts or transplants, or in *in vitro* culture.

Investigation of genetic function may utilize non-mammalian models, particularly using those organisms that are biologically and genetically well-characterized, such as *C. elegans*, *D. melanogaster* and *S. cerevisiae*. For example, transposon (Tc1) insertions in the nematode homolog of an Ngn3 gene or a promoter region of an Ngn3 gene may be made. The Ngn3 gene sequences may be used to knock-out or to complement defined genetic lesions in order to determine the physiological and biochemical pathways involved in function of islet cells. It is well known that human genes can complement mutations in lower eukaryotic models.

Production of Ngn3 Polypeptides

Ngn3-encoding nucleic acid may be employed to synthesize full-length Ngn3 polypeptides or fragments thereof, particularly fragments corresponding to functional domains; DNA binding sites; *etc.*; and including fusions of the subject polypeptides to other proteins or parts thereof. For expression, an expression cassette may be employed, providing for a transcriptional and translational initiation region, which may be inducible or constitutive, where the coding region is operably linked under the transcriptional control of the transcriptional initiation region, and a transcriptional and translational termination region. Various transcriptional initiation regions may be employed that are functional in the expression host.

As discussed above, the invention encompasses both isolated. naturally-occurring Ngn3 polypeptides, as well as recombinant Ngn3 polypeptides and functional equivalents of such recombinant and/or naturally-occurring Ngn3 polypeptides, *e.g.*, biologically active variants sharing substantial or significant amino acid sequence similarity and/or sequence identity with an Ngn3 amino acid sequence provided herein. Substantial identity, when referring to the Ngn3 polypeptides of the invention are polypeptides having at least about 70%, typically at least about 80% and preferably at least about 90% to about 95% identity to the amino acid sequence of SEQ ID NO: 2, or that are encoded by polynucleotides which will hybridize under stringent conditions to a polynucleotide having the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3.. Percent identity for the polypeptides of the invention is determined using the BLASTP program with the default settings as described at http://www.ncbi.nlm.nih.gov/ cgi-bin/BLAST/nph-newblast?Jform=0 with the DUST filter selected. The DUST filter is described at http://www.ncbi.nlm.nih.gov /BLAST/filtered.html.

Accordingly, the Ngn3 polynucleotides and polypeptides of this invention include, without limitation, Ngn3 polypeptides and polynucleotides found in primates, rodents, canines, felines, equines, nematodes, yeast and the like; and the natural and non-natural variants thereof.

The polypeptides may be expressed in prokaryotes or eukaryotes in accordance with conventional ways, depending upon the purpose for expression. For large scale production of the protein, a unicellular organism, such as *E. coli, B. subtilis, S. cerevisiae*, or cells of a higher organism such as vertebrates, particularly mammals, *e.g.* COS 7 cells, may be used as the expression host cells. In many situations, it may be desirable to express the Ngn3 genes in

mammalian cells, especially where the encoded polypeptides will benefit from native folding and post-translational modifications. Small peptides can also be synthesized in the laboratory.

With the availability of the polypeptides in large amounts, by employing an expression host, the polypeptides may be isolated and purified in accordance with conventional ways. A lysate may be prepared of the expression host and the lysate purified using HPLC, exclusion chromatography, gel electrophoresis, affinity chromatography, or other purification technique. The purified polypeptide will generally be at least about 80% pure, preferably at least about 90% pure, and may be up to and including 100% pure. Pure is intended to mean free of other proteins, as well as cellular debris.

The Ngn3 polypeptides can be used for the production of antibodies, where short fragments provide for antibodies specific for the particular polypeptide, and larger fragments or the entire protein allow for the production of antibodies over the surface of the polypeptide. Antibodies may be raised to the wild-type or variant forms of Ngn3. Antibodies may be raised to isolated peptides corresponding to these domains, or to the native protein, *e.g.* by immunization with cells expressing Ngn3, immunization with liposomes having Ngn3 polypeptides inserted in the membrane, *etc.*

Antibodies are prepared in accordance with conventional ways, where the expressed polypeptide or protein is used as an immunogen, by itself or conjugated to known immunogenic carriers, *e.g.* KLH, pre-S HBsAg, other viral or eukaryotic proteins, or the like. Various adjuvants may be employed, with a series of injections, as appropriate. For monoclonal antibodies, after one or more booster injections, the spleen is isolated, the lymphocytes immortalized by cell fusion, and then screened for high affinity antibody binding. The immortalized cells, *i.e.* hybridomas, producing the desired antibodies may then be expanded. For further description, see Monoclonal Antibodies: A Laboratory Manual, Harlow and Lane eds., Cold Spring Harbor Laboratories, Cold Spring Harbor, New York, 1988. If desired, the mRNA encoding the heavy and light chains may be isolated and mutagenized by cloning in *E. coli*, and the heavy and light chains mixed to further enhance the affinity of the antibody. Alternatives to in vivo immunization as a method of raising antibodies include binding to phage "display" libraries, usually in conjunction with in vitro affinity maturation.

Isolation of Ngn3 Allelic Variants and Homologues in Other Species

Other mammalian Ngn3 genes can be identified and their function characterized using the Ngn3 genes used in the present invention. Other Ngn3 genes of interest include, but are not limited to, mammalian (*e.g.*, human, rodent (e.g, murine, or rat), bovine, feline, canine, and the like) and non-mammalian (*e.g.*, chicken, reptile, and the like). Methods for identifying, isolating, sequencing, and characterizing an unknown gene based upon its homology to a known gene sequence are well known in the art (see, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, CSH Press 1989.

Drug Screening

The animal models of the invention, as well as methods using the Ngn3 polypeptides in vitro, can be used to identify candidate agents that affect Ngn3 expression (*e.g.*, by affecting Ngn3 promoter function) or that interact with Ngn3 polypeptides. Agents of interest can include those that enhance, inhibit, regulate, or otherwise affect Ngn3 activity and/or expression. Agents that alter Ngn3 activity and/or expression can be used to, for example, treat or study disorders associated with decreased Ngn3 activity (*e.g.*, diabetes or other pancreatic disorders), and/or to facilitate development of islet cell precursors either *in vitro* or *in vivo*. Candidate agents is meant to include synthetic molecules (e.g., small molecule drugs, peptides, or other synthetically produced molecules or compounds, as well as recombinantly produced gene products) as well as naturally-occurring compounds (e.g., polypeptides, endogenous factors present in insulin-producing, hormones, plant extracts, and the like).

Drug Screening Assays

Of particular interest in the present invention is the identification of agents that have activity in affecting Ngn3 expression and/or function. Such agents are candidates for development of treatments for, for example, diabetes or other condition that may be associated with altered Ngn3 activity. Drug screening identifies agents that provide a replacement or enhancement for Ngn3 function in affected cells. Conversely, agents that reverse or inhibit Ngn3 function may provide a means to regulate insulin production. Of particular interest are screening assays for agents that have a low toxicity for human cells.

The term "agent" as used herein describes any molecule, *e.g.* protein or pharmaceutical, with the capability of altering or mimicking the expression or physiological function of Ngn3. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, *i.e.* at zero concentration or below the level of detection.

Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including, but not limited to: peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, *etc.* to produce structural analogs.

Screening of Candidate Agents In Vivo

Agents can be screened for their ability to affect Ngn3. expression or function or to mitigate an undesirable phenotype (e.g., a symptom) associated with an alteration in Ngn3 expression or function. In a preferred embodiment, screening of candidate agents is performed in vivo in a transgenic animal described herein. Transgenic animals suitable for use

in screening assays include any transgenic animal having an alteration in Ngn3 expression, and can include transgenic animals having, for example, an exogenous and stably transmitted human Ngn3 gene sequence, a reporter gene composed of a (removed human) Ngn3 promoter sequence operably linked to a reporter gene (e.g,. β-galactosidase, CAT, or other gene that can be easily assayed for expression), or a homozygous or heterozygous knockout of an Ngn3 gene. The transgenic animals can be either homozygous or heterozygous for the genetic alteration and, where a sequence is introduced into the animal's genome for expression, may contain multiple copies of the introduced sequence. Where the *in vivo* screening assay is to identify agents that affect the activity of the Ngn3 promoter, the Ngn3 promoter can be operably linked to a reporter gene (*e.g.*, luciferase) and integrated into the non-human host animal's genome or integrated into the genome of a cultured mammalian cell line.

The candidate agent is administered to a non-human, transgenic animal having altered Ngn3 expression, and the effects of the candidate agent determined. The candidate agent can be administered in any manner desired and/or appropriate for delivery of the agent in order to effect a desired result. For example, the candidate agent can be administered by injection (e.g., by injection intravenously, intramuscularly, subcutaneously, or directly into the tissue in which the desired affect is to be achieved), orally, or by any other desirable means. Normally, the in vivo screen will involve a number of animals receiving varying amounts and concentrations of the candidate agent (from no agent to an amount of agent hat approaches an upper limit of the amount that can be delivered successfully to the animal), and may include delivery of the agent in different formulation. The agents can be administered singly or can be combined in combinations of two or more, especially where administration of a combination of agents may result in a synergistic effect.

The effect of agent administration upon the transgenic animal can be monitored by assessing Ngn3 function as appropriate (e.g., by examining expression of a reporter or fusion gene), or by assessing a phenotype associated with the Ngn3 expression. For example, where the transgenic animal used in the screen contains a defect in Ngn3 expression (e.g., due to a knock-out of the gene), the effect of the candidate agent can be assessed by determining levels of hormones produced in the mouse relative to the levels produced in the Ngn3 defective transgenic mouse and/or in wildtype mice (e.g, by assessing levels of insulin). Methods for assaying insulin are well known in the art. Where the *in vivo* screening assay is

to identify agents that affect the activity of the Ngn3 promoter and the non-human transgenic animal (or cultured mammalian cell line) comprises an Ngn3 promoter operably linked to a reporter gene, the effects of candidate agents upon Ngn3 promoter activity can be screened by, for example, monitoring the expression from the Ngn3 promoter (through detection of the reporter gene) and correlation of altered Ngn3 promoter activity with islet cell formation. Alternatively or in addition, Ngn3 promoter activity can be assessed by detection (qualitative or quantitative) of Ngn3 mRNA or protein levels. Where the candidate agent affects Ngn3 expression, and/or affects an Ngn3-associated phenotype, in a desired manner, the candidate agent is identified as an agent suitable for use in therapy of an Ngn3-associated disorder and/or to facilitate development of islet precursor cells to mature β-cells either *in vivo* or *in vitro*.

Screening of Candidate Agents In Vitro

In addition to screening of agents in Ngn3 transgenic animals, a wide variety of in vitro assays may be used for this purpose, including labeled in vitro protein-protein binding assays, protein-DNA binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, and the like. For example, by providing for the production of large amounts of Ngn3 protein, one can identify ligands or substrates that bind to, modulate or mimic the action of the proteins. The purified protein may also be used for determination of three-dimensional crystal structure, which can be used for modeling intermolecular interactions, transcriptional regulation, *etc.*

The screening assay can be a binding assay, wherein one or more of the molecules may be joined to a label, and the label directly or indirectly provide a detectable signal. Various labels include radioisotopes, fluorescers, chemiluminescers, enzymes, specific binding molecules, particles, *e.g.* magnetic particles, and the like. Specific binding molecules include pairs, such as biotin and streptavidin, digoxin and antidigoxin etc. For the specific binding members, the complementary member would normally be labeled with a molecule that provides for detection, in accordance with known procedures.

A variety of other reagents may be included in the screening assays described herein. Where the assay is a binding assay, these include reagents like salts, neutral proteins, e.g. albumin, detergents, etc that are used to facilitate optimal protein-protein binding, protein-DNA binding, and/or reduce non-specific or background interactions. Reagents that improve

the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, *etc.* may be used. The mixture of components are added in any order that provides for the requisite binding. Incubations are performed at any suitable temperature, typically between 4 and 40°C. Incubation periods are selected for optimum activity, but may also be optimized to facilitate rapid high-throughput screening. Typically between 0.1 and 1 hours will be sufficient.

Other assays of interest detect agents that mimic Ngn3 function. For example, candidate agents are added to a cell that lacks functional Ngn3, and screened for the ability to reproduce Ngn3 activity in a functional assay.

Many mammalian genes have homologs in yeast and lower animals. The study of such homologs' physiological role and interactions with other proteins in vivo or in vitro can facilitate understanding of biological function. In addition to model systems based on genetic complementation, yeast has been shown to be a powerful tool for studying protein-protein interactions through the two hybrid system described in Chien et al. 1991 Proc. Natl. Acad. Sci. USA 88:9578-9582. Two-hybrid system analysis is of particular interest for exploring transcriptional activation by Ngn3 proteins and to identify cDNAs encoding polypeptides that interact with Ngn3.

Identified Candidate Agents

The compounds having the desired pharmacological activity may be administered in a physiologically acceptable carrier to a host for treatment of a condition attributable to a defect in Ngn3 function (*e.g.*, a disorder associated with reduced insulin levels (e.g., diabetes (Type 1 or Type 2 diabetes, particularly Type 1 diabetes)). The compounds may also be used to enhance Ngn3 function. The therapeutic agents may be administered in a variety of ways, orally, topically, parenterally *e.g.* subcutaneously, intraperitoneally, by viral infection, intravascularly, *etc.* Inhaled treatments are of particular interest. Depending upon the manner of introduction, the compounds may be formulated in a variety of ways. The concentration of therapeutically active compound in the formulation may vary from about 0.1-100 wt.%.

The pharmaceutical compositions can be prepared in various forms, such as granules, tablets, pills, suppositories, capsules, suspensions, salves, lotions and the like. Pharmaceutical grade organic or inorganic carriers and/or diluents suitable for oral and topical use can be

used to make up compositions containing the therapeutically-active compounds. Diluents known to the art include aqueous media, vegetable and animal oils and fats. Stabilizing Agents, wetting and emulsifying Agents, salts for varying the osmotic pressure or buffers for securing an adequate pH value, and skin penetration enhancers can be used as auxiliary agents.

Pharmacogenetics

Pharmacogenetics is the linkage between an individual's genotype and that individual's ability to metabolize or react to a therapeutic agent. Differences in metabolism or target sensitivity can lead to severe toxicity or therapeutic failure by altering the relation between bioactive dose and blood concentration of the drug. In the past few years, numerous studies have established good relationships between polymorphisms in metabolic enzymes or drug targets, and both response and toxicity. These relationships can be used to individualize therapeutic dose administration.

Genotyping of polymorphic alleles is used to evaluate whether an individual will respond well to a particular therapeutic regimen. The polymorphic sequences are also used in drug screening assays, to determine the dose and specificity of a candidate therapeutic agent. A candidate Ngn3 polymorphism is screened with a target therapy to determine whether there is an influence on the effectiveness in treating, for example, diabetes. Drug screening assays are performed as described above. Typically two or more different sequence polymorphisms are tested for response to a therapy. Therapies for diabetes currently include replacement therapy via administration of insulin and administration of drugs that increase insulin secretion (sulfonylureas) and drugs that reduce insulin resistance (such as troglitazone).

Where a particular sequence polymorphism correlates with differential drug effectiveness, diagnostic screening may be performed. Diagnostic methods have been described in detail in a preceding section. The presence of a particular polymorphism is detected, and used to develop an effective therapeutic strategy for the affected individual.

Detection of Ngn3 Associated Disorders

Diagnosis of Ngn3-associated disorders is performed by protein, DNA or RNA sequence and/or hybridization analysis of any convenient sample from a patient, e.g. biopsy material, blood sample, scrapings from cheek, etc. A nucleic acid sample from a patient

having a disorder that may be associated with Ngn3, is analyzed for the presence of a predisposing polymorphism in Ngn3. A typical patient genotype will have at least one predisposing mutation on at least one chromosome. The presence of a polymorphic Ngn3 sequence that affects the activity or expression of the gene product, and confers an increased susceptibility to an Ngn3 associated disorder (e.g, hyperglycemia, diabetes, and the like) is considered a predisposing polymorphism. Individuals are screened by analyzing their DNA or mRNA for the presence of a predisposing polymorphism, as compared to sequence from an unaffected individual(s). Specific sequences of interest include, for example, any polymorphism that is associated with a diabetic syndrome, especially with Type 2 diabetes, or is otherwise associated with diabetes, including, but not limited to, insertions, substitutions and deletions in the coding region sequence, intron sequences that affect splicing, or promoter or enhancer sequences that affect the activity and expression of the protein.

Screening may also be based on the functional or antigenic characteristics of the protein. Immunoassays designed to detect predisposing polymorphisms in Ngn3 proteins may be used in screening. Where many diverse mutations lead to a particular disease phenotype, functional protein assays can be effective screening tools.

Biochemical studies may be performed to determine whether a candidate sequence polymorphism in the Ngn3 coding region or control regions is associated with disease. For example, a change in the promoter or enhancer sequence that affects expression of Ngn3 may result in predisposition to diabetes. Expression levels of a candidate variant allele are compared to expression levels of the normal allele by various methods known in the art. Methods for determining promoter or enhancer strength include quantitation of the expressed natural protein; insertion of the variant control element into a vector with a reporter gene such as β-galactosidase, luciferase, chloramphenicol acetyltransferase, etc. that provides for convenient quantitation; and the like. The activity of the encoded Ngn3 protein may be determined by comparison with the wild-type protein.

A number of methods are available for analyzing nucleic acids for the presence of a specific sequence. Where large amounts of DNA are available, genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. Cells that express Ngn3 genes, such as pancreatic cells, may be used as a source of mRNA, which may be assayed directly or reverse transcribed into cDNA for analysis. The nucleic acid may be amplified by conventional techniques, such as

the polymerase chain reaction (PCR), to provide sufficient amounts for analysis. The use of the polymerase chain reaction is described in Saiki, et al. 1985 Science 239:487; a review of current techniques may be found in Sambrook, et al. Molecular Cloning: A Laboratory Manual. CSH Press 1989, pp.14.2–14.33. Amplification may also be used to determine whether a polymorphism is present, by using a primer that is specific for the polymorphism. Alternatively, various methods are known in the art that utilize oligonucleotide ligation as a means of detecting polymorphisms, for examples see Riley et al. 1990 Nucl. Acid Res. 18:2887-2890; and Delahunty et al. 1996 Am. J. Hum. Genet. 58:1239-1246.

A detectable label may be included in an amplification reaction. Suitable labels include fluorochromes, e.g. fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, e.g. ^{32}P, ^{35}S, ^{3}H; etc. The label may be a two stage system, where the amplified DNA is conjugated to biotin, haptens, etc. having a high affinity binding partner, e.g. avidin, specific antibodies, etc., where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

The sample nucleic acid, e.g. amplified or cloned fragment, is analyzed by one of a number of methods known in the art. The nucleic acid may be sequenced by dideoxy or other methods, and the sequence of bases compared to either a neutral Ngn3 sequence (*e.g.*, an Ngn3 sequence from an unaffected individual). Hybridization with the variant sequence may also be used to determine its presence, by Southern blots, dot blots, etc. The hybridization pattern of a control and variant sequence to an array of oligonucleotide probes immobilized on a solid support, as described in US 5,445,934, or in WO95/35505, may also be used as a means of detecting the presence of variant sequences. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), mismatch cleavage detection, and heteroduplex analysis in gel matrices are used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility. Alternatively, where a polymorphism creates or destroys a recognition site for a restriction endonuclease (restriction fragment length polymorphism, RFLP), the sample is

digested with that endonuclease, and the products size fractionated to determine whether the fragment was digested. Fractionation is performed by gel or capillary electrophoresis, particularly acrylamide or agarose gels.

The hybridization pattern of a control and variant sequence to an array of oligonucleotide probes immobilized on a solid support, as described in US 5,445,934, or in WO95/35505, may be used as a means of detecting the presence of variant sequences. In one embodiment of the invention, an array of oligonucleotides are provided, where discrete positions on the array are complementary to at least a portion of mRNA or genomic DNA of the Ngn3 locus. Such an array may comprise a series of oligonucleotides, each of which can specifically hybridize to a nucleic acid sequence, *e.g.*, mRNA, cDNA, genomic DNA, *etc.* from the Ngn3 locus. Usually such an array will include at least 2 different polymorphic sequences, i.e. polymorphisms located at unique positions within the locus, usually at least about 5, more usually at least about 10, and may include as many as 50 to 100 different polymorphisms. The oligonucleotide sequence on the array will usually be at least about 12 nt in length, may be the length of the provided polymorphic sequences, or may extend into the flanking regions to generate fragments of 100 to 200 nt in length. For examples of arrays, see Hacia et al. 1996 Nature Genetics 14:441-447; Lockhart et al. 1996 Nature Biotechnol. 14:1675-1680; and De Risi et al. 1996 Nature Genetics 14:457-460.

Antibodies specific for Ngn3 polymorphisms may be used in screening immunoassays. A reduction or increase in Ngn3 and/or presence of an Ngn3 disorder associated polymorphism is indicative that the suspected disorder is Ngn3-associated. A sample is taken from a patient suspected of having an Ngn3-associated disorder. Samples, as used herein, include tissue biopsies, biological fluids, organ or tissue culture derived fluids, and fluids extracted from physiological tissues, as well as derivatives and fractions of such fluids. The number of cells in a sample will generally be at least about 10^3, usually at least 10^4 more usually at least about 10^5. The cells may be dissociated, in the case of solid tissues, or tissue sections may be analyzed. Alternatively a lysate of the cells may be prepared.

Diagnosis may be performed by a number of methods. The different methods all determine the absence or presence or altered amounts of normal or abnormal Ngn3 in patient cells suspected of having a predisposing polymorphism in Ngn3. For example, detection may utilize staining of cells or histological sections, performed in accordance with conventional methods. The antibodies of interest are added to the cell sample, and incubated for a period

of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody may be labeled with radioisotopes, enzymes, fluorescers, chemiluminescers, or other labels for direct detection. Alternatively, a second stage antibody or reagent is used to amplify the signal. Such reagents are well known in the art. For example, the primary antibody may be conjugated to biotin, with horseradish peroxidase-conjugated avidin added as a second stage reagent. Final detection uses a substrate that undergoes a color change in the presence of the peroxidase. The absence or presence of antibody binding may be determined by various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, etc.

An alternative method for diagnosis depends on the in vitro detection of binding between antibodies and Ngn3 in a lysate. Measuring the concentration of Ngn3 binding in a sample or fraction thereof may be accomplished by a variety of specific assays. A conventional sandwich type assay may be used. For example, a sandwich assay may first attach Ngn3-specific antibodies to an insoluble surface or support. The particular manner of binding is not crucial so long as it is compatible with the reagents and overall methods of the invention. They may be bound to the plates covalently or non-covalently, preferably non-covalently.

The insoluble supports may be any compositions to which polypeptides can be bound, which is readily separated from soluble material, and which is otherwise compatible with the overall method. The surface of such supports may be solid or porous and of any convenient shape. Examples of suitable insoluble supports to which the receptor is bound include beads, e.g. magnetic beads, membranes and microtiter plates. These are typically made of glass, plastic (e.g. polystyrene), polysaccharides, nylon or nitrocellulose. Microtiter plates are especially convenient because a large number of assays can be carried out simultaneously, using small amounts of reagents and samples.

Patient sample lysates are then added to separately assayable supports (for example, separate wells of a microtiter plate) containing antibodies. Preferably, a series of standards, containing known concentrations of normal and/or abnormal Ngn3 is assayed in parallel with the samples or aliquots thereof to serve as controls. Preferably, each sample and standard will be added to multiple wells so that mean values can be obtained for each. The incubation time should be sufficient for binding, generally, from about 0.1 to 3 hr is sufficient. After incubation, the insoluble support is generally washed of non-bound components. Generally, a

dilute non-ionic detergent medium at an appropriate pH, generally 7-8, is used as a wash medium. From one to six washes may be employed, with sufficient volume to thoroughly wash non-specifically bound proteins present in the sample.

After washing, a solution containing a second antibody is applied. The antibody will bind Ngn3 with sufficient specificity such that it can be distinguished from other components present. The second antibodies may be labeled to facilitate direct, or indirect quantification of binding. Examples of labels that permit direct measurement of second receptor binding include radiolabels, such as ^{3}H or ^{125}I, fluorescers, dyes, beads, chemiluminescers, colloidal particles, and the like. Examples of labels which permit indirect measurement of binding include enzymes where the substrate may provide for a colored or fluorescent product. In a preferred embodiment, the antibodies are labeled with a covalently bound enzyme capable of providing a detectable product signal after addition of suitable substrate. Examples of suitable enzymes for use in conjugates include horseradish peroxidase, alkaline phosphatase, malate dehydrogenase and the like. Where not commercially available, such antibody-enzyme conjugates are readily produced by techniques known to those skilled in the art. The incubation time should be sufficient for the labeled ligand to bind available molecules. Generally, from about 0.1 to 3 hr is sufficient, usually 1 hr sufficing.

After the second binding step, the insoluble support is again washed free of non-specifically bound material. The signal produced by the bound conjugate is detected by conventional means. Where an enzyme conjugate is used, an appropriate enzyme substrate is provided so a detectable product is formed.

Other immunoassays are known in the art and may find use as diagnostics. Ouchterlony plates provide a simple determination of antibody binding. Western blots may be performed on protein gels or protein spots on filters, using a detection system specific for Ngn3 as desired, conveniently using a labeling method as described for the sandwich assay.

Other diagnostic assays of interest are based on the functional properties of Ngn3 proteins. Such assays are particularly useful where a large number of different sequence changes lead to a common phenotype. For example, a functional assay may be based on the transcriptional changes mediated by Ngn3 gene products. Other assays may, for example, detect conformational changes, size changes resulting from insertions, deletions or truncations, or changes in the subcellular localization of Ngn3 proteins.

In a protein truncation test, PCR fragments amplified from the Ngn3 gene or its transcript are used as templates for in vivo transcription/translation reactions to generate protein products. Separation by gel electrophoresis is performed to determine whether the polymorphic gene encodes a truncated protein, where truncations may be associated with a loss of function.

Diagnostic screening may also be performed for polymorphisms that are genetically linked to a predisposition for diabetes, particularly through the use of microsatellite markers or single nucleotide polymorphisms. Frequently the microsatellite polymorphism itself is not phenotypically expressed, but is linked to sequences that result in a disease predisposition. However, in some cases the microsatellite sequence itself may affect gene expression. Microsatellite linkage analysis may be performed alone, or in combination with direct detection of polymorphisms, as described above. The use of microsatellite markers for genotyping is well documented. For examples, see Mansfield et al. 1994 Genomics 24:225-233; Ziegle et al. 1992 Genomics 14:1026-1031; Dib et al., supra.

Microsatellite loci that are useful in the subject methods have the general formula:

$$U\ (R)_n\ U',\ \text{where}$$

U and U' are non-repetitive flanking sequences that uniquely identify the particular locus, R is a repeat motif, and n is the number of repeats. The repeat motif is at least 2 nucleotides in length, up to 7, usually 2-4 nucleotides in length. Repeats can be simple or complex. The flanking sequences U and U' uniquely identify the microsatellite locus within the human genome. U and U' are at least about 18 nucleotides in length, and may extend several hundred bases up to about 1 kb on either side of the repeat. Within U and U', sequences are selected for amplification primers. The exact composition of the primer sequences are not critical to the invention, but they must hybridize to the flanking sequences U and U', respectively, under stringent conditions. Criteria for selection of amplification primers are as previously discussed. To maximize the resolution of size differences at the locus, it is preferable to chose a primer sequence that is close to the repeat sequence, such that the total amplification product is between 100-500 nucleotides in length.

The number of repeats at a specific locus, n, is polymorphic in a population, thereby generating individual differences in the length of DNA that lies between the amplification primers. The number will vary from at least 1 repeat to as many as about 100 repeats or more.

The primers are used to amplify the region of genomic DNA that contains the repeats. Conveniently, a detectable label will be included in the amplification reaction, as previously described. Multiplex amplification may be performed in which several sets of primers are combined in the same reaction tube. This is particularly advantageous when limited amounts of sample DNA are available for analysis. Conveniently, each of the sets of primers is labeled with a different fluorochrome.

After amplification, the products are size fractionated. Fractionation may be performed by gel electrophoresis, particularly denaturing acrylamide or agarose gels. A convenient system uses denaturing polyacrylamide gels in combination with an automated DNA sequencer, see Hunkapillar et al. 1991 Science 254:59-74. The automated sequencer is particularly useful with multiplex amplification or pooled products of separate PCR reactions. Capillary electrophoresis may also be used for fractionation. A review of capillary electrophoresis may be found in Landers. et al. 1993 BioTechniques 14:98-111. The size of the amplification product is proportional to the number of repeats (n) that are present at the locus specified by the primers. The size will be polymorphic in the population, and is therefore an allelic marker for that locus.

Therapeutic Uses of Ngn3-Encoding Nucleic Acid

Ngn3-encoding nucleic acid can be introduced into a cell to accomplish transformation of the cell, preferably stable transformation, and the transformed cell subsequently implanted into a subject having a disorder characterized by a deficiency in insulin (*e.g.*, an Ngn3-associated disorder), depending upon the tissue into which the transformed cell is implanted. Preferably, the host cell to be transformed and implanted in the subject is derived from the individual who will receive the transplant (e.g., to provide an autologous transplant). Where the transformed cells are to be inserted into individual (*e.g.*, into the pancreas, liver, abdominal cavity, etc.), the cells into which the nucleic acid is introduced are preferably stem cells capable of developing into β cells within the pancreatic tissue environment, e.g., stem cells derived from pancreatic tissue, gastrointestinal tissue, or cells capable of expression of insulin upon expression of the Ngn3-encoding nucleic acid.

For example, in a subject having Type I diabetes, gastrointestinal stem cells can be isolated from the affected subject, the cells transformed with Ngn3-encoding DNA, and the transformed cells implanted in the affected subject to provide for insulin production, or the

transformed cells cultured so as to facilitate development of the cells into insulin-producing β-cells.

Introduction of the Ngn3-encoding nucleic acid into the cell can be accomplished according to methods well known in the art (e.g., through use of electroporation, microinjection, lipofection infection with a recombinant (preferably replication-deficient) virus, and other means well known in the art). Preferably, the Ngn3-encoding nucleic acid is operably linked to a promoter that facilitates a desired level of Ngn3 polypeptide expression (e.g., a promoter derived from CMV, SV40, adenovirus, or a tissue-specific or cell type-specific promoter). Transformed cells containing the Ngn3-encoding nucleic acid can be selected and/or enriched via, for example, expression of a selectable marker gene present in the Ngn3-encoding construct or that is present on a plasmid that is co-transfected with the Ngn3-encoding construct. Typically selectable markers provide for resistance to antibiotics such as tetracycline, hygromycin, neomycin, and the like. Other markers can include thymidine kinase and the like.

The ability of the transformed cells to express the Ngn3-encoding nucleic acid can be assessed by various methods known in the art. For example, Ngn3 expression can be examined by Northern blot to detect mRNA which hybridizes with a DNA probe derived from the relevant gene. Those cells that express the desired gene can be further isolated and expanded in in vitro culture using methods well known in the art. The host cells selected for transformation with Ngn3-encoding DNA will vary with the purpose of the *ex vivo* therapy (e.g., insulin production), the site of implantation of the cells, and other factors that will vary with a variety of factors that will be appreciated by the ordinarily skilled artisan.

Methods for engineering a host cell for expression of a desired gene product(s) and implantation or transplantion of the engineered cells (e.g., ex vivo therapy) are known in the art (see, e.g., Gilbert et al. 1993 "Cell transplantation of genetically altered cells on biodegradable polymer scaffolds in syngeneic rats," Transplantation 56:423-427). For expression of a desired gene in exogenous or autologous cells and implantation of the cells (e.g., islet cells) into pancreas, see, e.g., Docherty 1997 "Gene therapy for diabetes mellitus," Clin Sci (Colch) 92:321-330; Hegre et al. 1976 "Transplantation of islet tissue in the rat," Acta Endocrinol Suppl (Copenh) 205:257-281; Sandler et al. 1997 "Assessment of insulin secretion in vitro from microencapsulated fetal porcine islet-like cell clusters and rat, mouse, and human pancreatic islets," Transplantation 63:1712-1718; Calafiore 1997 "Perspectives in

pancreatic and islet cell transplantation for the therapy of IDDM," Diabetes Care 20:889-896; Kenyon et al. 1996 "Islet cell transplantation: beyond the paradigms," Diabetes Metab Rev 12:361-372; Sandler; Chick et al. 1977 Science "Artificial pancreas using living beta cells:. effects on glucose homeostasis in diabetic rats," 197:780-782.

After expansion of the transformed cells *in vitro*, the cells are implanted into the mammalian subject, preferably into the tissue from which the cells were originally derived, by methods well known in the art. The number of cells implanted is a number of cells sufficient to provide for expression of levels of Ngn3 sufficient to provide for enhanced levels of insulin. The number cells to be transplanted can be determined based upon such factors as the levels of polypeptide expression achieved *in vitro*, and/or the number of cells that survive implantation. Preferably the cells are implanted in an area of dense vascularization, and in a manner that minimizes evidence of surgery in the subject. The engraftment of the implant of transformed cells is monitored by examining the mammalian subject for classic signs of graft rejection, i.e., inflammation and/or exfoliation at the site of implantation, and fever.

Alternatively, Ngn3-encoding nucleic acid can be delivered directly to an affected subject to provide for Ngn3 expression in a target cell (e.g., a pancreatic cell, gut cell, liver cell, or other organ cell capable of expressing Ngn3 and providing production of insulin), thereby promoting development of the cell into an insulin-producing cell (e.g., in pancreas) or to cure a defect in Ngn3 expression in the subject. Methods for *in vivo* delivery of a nucleic acid of interest for expression in a target cell are known in the art. For example, in vivo methods of gene delivery normally employ either a biological means of introducing the DNA into the target cells (e.g., a virus containing the DNA of interest) or a mechanical means to introduce the DNA into the target cells (e.g., direct injection of DNA into the cells, liposome fusion, pneumatic injection using a "gene gun," or introduction of the DNA via a duct of the pancreas). For other methods of introduction of a DNA of interest into a cell in vivo, also see Bartlett et al. 1997 "Use of biolistic particle accelerator to introduce genes into isolated islets of Langerhans," Transplant Proc 29:2201-2202; Furth 1997 "Gene transfer by biolistic process," Mol Biotechnol 7:139-143; Gainer et al. 1996 "Successful biolistic transformation of mouse pancreatic islets while preserving cellular function," Transplantation 61:1567-1571; Docherty 1997 "Gene therapy for diabetes mellitus," Clin Sci (Colch) 92:321-330; Maeda et al. 1994 "Gastroenterology 1994 "Adenovirus-mediated transfer of human lipase complementary DNA to the gallbladder," 106:1638-1644.

The amount of DNA and/or the number of infectious viral particles effective to infect the targeted tissue, transform a sufficient number of cells, and provide for production of a desired level of insulin can be readily determined based upon such factors as the efficiency of the transformation *in vitro* and the susceptibility of the targeted secretory gland cells to transformation. For example, the amount of DNA injected into the pancreas of a human is, for example, generally from about 1 μg to 750 mg, preferably from about 500 μg to 500 mg, more preferably from about 10 mg to 200 mg, most preferably about 100 mg. Generally, the amounts of DNA can be extrapolated from the amounts of DNA effective for delivery and expression of the desired gene in an animal model. For example, the amount of DNA for delivery in a human is roughly 100 times the amount of DNA effective in a rat.

Regardless of whether the Ngn3-encoding DNA is introduced *in vivo* or *ex vivo*, the DNA (or cells expressing the DNA) can be administered in combination with other genes and other agents. In addition, Ngn3-encoding DNA (or recombinant cells expressing Ngn3 DNA) can be used therapeutically for disorders associated with, for example, a decrease in insulin production, but which are not associated with an alteration in Ngn3 function per se. For example, an increase in Ngn3 may cause an increase in the number of mature β cells, and thus an increase in insulin production, in an individual that has decreased insulin production from some other cause not related to function of Ngn3.

Identification of Islet Cell Precursors and Development of β-Cells Using Ngn3

As described in more detail in the Examples below, the temporal and spatial pattern of Ngn3 expression indicates that Ngn3 can be used as a marker for islet cell precursors. This feature of Ngn3 expression can be exploited to provide compositions and methods to identify and isolate islet cell precursors. For example, pancreatic tissue can be obtained from a subject, and a single cell suspension obtained from the tissue. The single cell cultures can then be expanded in culture, and representative cells from the single cell cultures analyzed for Ngn3 expression. Ngn3 expression can be analyzed by, for example, detection of Ngn3-encoding mRNA (*e.g.*, by PCR amplification using a probe derived from an Ngn3-encoding sequence) or by detection of the Ngn3 polypeptide in cell lysates using an anti-Ngn3 antibody. Cells that express Ngn3 are identified as being islet cell precursors. The cells of the corresponding culture could then be expanded and/or used to derive mature β-cells in culture,

and the mature β-cells implanted into the subject, *e.g.*, either into the same subject from whom the cells were initially obtained or into a different subject.

Ngn3 is also useful for monitoring development of islet cell precursors into mature β-cells. In short, Ngn3 expression can be monitored in an *in vitro* culture to determine when the cells become mature β-cells. For example, cells that express Ngn3 are at an earlier stage of β-cell development. Once Ngn3 expression decreases or becomes substantially undetectable, the cell can be identified as having developed into a mature β-cell. The cells can be screened for other markers of islet cell development, as well as for insulin production.

EXAMPLES

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to carry out the invention and is not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

Example 1: Detection of Ngn3 Expression in Murine Pancreas

Members of the basic helix-loop-helix (bHLH) family of transcription factors regulate growth and differentiation of numerous cell types. Insulin gene expression is activated by a heterodimeric complex of two bHLH proteins: a ubiquitously expressed (class A) protein and a cell-type-specific (class B) partner, BETA2/neuroD1. BETA2/neuroD1 is also important for β-cell development. The targeted disruption of the BETA2/neuroD1 gene in mice leads to a marked reduction of the β-cell mass at birth due to increased apoptosis of islet cells late in fetal development. There is no apparent defect, however, in β-cell formation or insulin gene expression, despite the postulated importance of this factor in β-cell differentiation.

Assuming that this modest phenotype reflected the redundant expression of closely related class B bHLH proteins in the endocrine pancreas, the inventors searched for additional members of the family by reverse transcriptase-polymerase chain reaction (RT-PCR) using degenerate oligonucleotides primers based on conserved amino acid sequences in the bHLH domain of the class B bHLH proteins (Sommer *et al.* 1996 *Mol. Cell. Neurosci.* 8:221). PCR

analysis revealed that pancreatic endocrine cell lines and isolated adult islets not only express neuroD1, but also several other members of the family of neural class B bHLH genes as well, including mash1, neuroD2 and 4 and neurogenins (ngn) 1, 2 and 3. This remarkable degree of redundancy could compensate for the loss of BETA2/neuroD1 in mice. The two most commonly amplified sequences encoded neuroD4 and Ngn3, but *in situ* hybridization studies in mouse pancreas showed highest expression of neuroD1 and Ngn3. These results were confirmed by immunohistochemistry.

Ngn3 is detected earliest at embryonic day 11.5 (e11.5) in the mouse, increases to a maximum at e15.5 and decreases at e18.5, with no staining seen in the adult pancreas. Ngn3 is detected in the nuclei of scattered ductal cells and periductal cells, and there was no co-staining with any of the four islet hormones (insulin, glucagon, somatostatin and pancreatic polypeptide). This temporal and spatial pattern of expression implicated *Ngn3* as a marker for islet cell precursors. Nkx6.1, a specific marker for future beta-cells, was expressed in 10-20% of the Ngn3 positive cells, further supporting the use of Ngn3 as a marker for islet cell precursors. The peak of *Ngn3* expression at e15.5 also corresponds with the peak of new beta-cell formation in the fetus. Our data supports a model in which *Ngn3* acts upstream of BETA2/neuroD1 and other islet differentiation factors, marking islet cell precursors, but switching off prior to final differentiation.

Example 2: Isolation and Sequencing of a Human Ngn3 Polypeptide-Encoding Polynucleotide

A probe derived from a cloned fragment of the murine Ngn3 gene (Sommer *et al.*, *supra*) was used to screen a human genomic library. This screen resulted in the isolation of the genomic sequence provided as SEQ ID NO:1 in the sequence listing. Based on mapping of the murine start site using 5' RACE of mouse fetal pancreatic RNA, the transcriptional start site in the human Ngn3-encoding sequence is at nucleotide residue 2643. The coding sequence is between nucleotide residues 3022-3663, with a stop site at 3664-3666. No introns are within the 5' untranslated region (UTR) or the coding sequence of SEQ ID NO:1.

The promoter of Ngn3 is of interest, particularly given that is it exceptionally well-conserved between mouse, rat, and human. Given the role of Ngn3 in pancreatic and islet cell development, the Ngn3 promoter is likely key to determining the number of islet cells in the mature pancreas. The regulatory region corresponding to the human Ngn3 promoter

comprises sequences up to approximately 500 bp upstream of the transcription start site within the human Ngn3 promoter (*e.g.*, from about 2144 to the transcriptional start site at 2643).

FISH was used to identify the location of Ngn3 on the human chromosome at 10q22.1-22.2.

Example 3: Isolation and Sequencing of a Murine Ngn3 Polypeptide-Encoding Polynucleotide and Promoter

The full-length murine Ngn3 sequence and its 5' flanking sequences, which included the murine Ngn3 promoter, were obtained by sequencing a previously obtained mouse genomic DNA fragment (Sommer, *et al.*, *supra*). The murine Ngn3 sequence is provided in the Sequence Listing as SEQ ID NO:3, with the encoded polypeptide provided as SEQ ID NO:4. The transcriptional start site was determined using the 5' RACE method and confirmed using Rnase protection with RNA from fetal mouse pancreas, and iis at nucleotide residue 719; the coding sequence for murine Ngn3 begins at nucleotide residue 1093. The promoter comprises a region approximately 500 bp upstream of the transcription start site.

The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

CLAIMS

What is claimed is:

1. An isolated human neurogenin3 (Ngn3) polypeptide.

2. The human Ngn3 polypeptide of claim 1, comprising an amino acid sequence of SEQ ID NO:2.

3. The human Ngn3 polypeptide of claim 1 comprising an amino acid sequence having at least about 70% amino acid sequence identity with the amino acid sequence of SEQ ID NO:2.

4 An isolated polynucleotide sequence or complement thereof comprising a polynucleotide sequence encoding a human Ngn3 polypeptide of claim 1.

5. The isolated polynucleotide of claim 4, wherein the Ngn3 polypeptide has an amino acid sequence that is substantially identical to the amino acid sequence of SEQ ID NO:2.

6. The isolated polynucleotide sequence of claim 4 comprising a polynucleotide sequence of nucleotides 3022-3063 of SEQ ID NO:1.

7. An isolated polynucleotide sequence that hybridizes under stringent conditions to the polynucleotide sequence of nucleotides 3022-3063 of SEQ ID NO:1.

8. A recombinant expression vector comprising the polynucleotide sequence of claim 4.

9. An isolated recombinant host cell comprising a polynucleotide sequence encoding the polypeptide of claim 1.

10. A method for producing the human Ngn3 polypeptide of claim 1, the method comprising the steps of:

a) culturing a recombinant host cell containing a human Ngn3 polypeptide-encoding polynucleotide sequence under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

11. An isolated antibody that specifically binds a human Ngn3 polypeptide of claim 1.

12. A method for identifying a polynucleotide homologous to the polynucleotide of claim 4. the method comprising the steps of:

contacting a polynucleotide probe with a test polynucleotide, the probe comprising at least 15 contiguous nucleotides of a polynucleotide sequence encoding a human Ngn3 polypeptide; and

detecting hybridization of the probe with the test polynucleotide;

wherein detection of hybridization of the probe to the test polynucleotide indicates that the polynucleotide shares sequence homology with the human Ngn3 polypeptide-encoding polynucleotide.

13. A method for identifying an islet cell precursor, the method comprising the step of analyzing a cell for expression of an neurogenin3 (Ngn3) gene product, wherein detection of the Ngn3 gene product is indicative of an islet cell precursor.

14. An isolated nucleic acid sequence comprising a neurogenin3 (Ngn3) promoter.

15. The isolated nucleic acid sequence of claim 14, wherein the Ngn3 promoter is a human neurogenin 3 promoter.

16. The isolated nucleic acid sequence of claim 14, wherein the sequence comprises a nucleotide sequence of a region 5' of nucleotide residue 2643 of SEQ ID NO:1.

17. The isolated nucleic acid sequence of claim 14, wherein the Ngn3 promoter is a murine neurogenin3 promoter.

18. The isolated nucleic acid sequence of claim 17, wherein the Ngn3 promoter comprises a nucleotide sequence of a region 5' of nucleotide residue 719 of SEQ ID NO:3.

19. A method for identifying a biologically active agent that modulates human neurogenin3 (Ngn3) activity, the method comprising:

combining a candidate agent with any one of:

(a) a human Ngn3 polypeptide;

(b) a recombinant cell comprising a nucleic acid encoding a human Ngn3 polypeptide; or

(c) a recombinant cell comprising a nucleic acid encoding a mammalian Ngn3 promoter sequence operably linked to a nucleic acid encoding a report gene;and

determining the effect of said agent on Ngn3 activity.

20. A method for detecting in a subject a predisposition to a defect in pancreatic islet cells function or formation associated with a defect in neurogenin3 (Ngn3) activity, the method comprising:

analyzing the genomic DNA or mRNA of an individual for the presence of at least one predisposing alteration in a genomic Ngn3 sequence;

wherein the presence of the altered genomic Ngn3 sequence is indicative of an increased susceptibility to a defect in pancreatic islet cell function or formation.

21. The method of claim 20, wherein the alteration is in an Ngn3 promoter sequence.

22. The method of claim 20, wherein the alteration is in a genomic sequence encoding an Ngn3 polypeptide.

23. A method for producing a human pancreatic islet cell, the method comprising;

identifying a human pancreatic islet cell precursor by detection of expression of human neurogenin3 (Ngn3); and

expanding the identified precursor cell in vitro;

wherein expansion of the identified cells produces a human pancreatic islet cell.

24. A pancreatic islet cell produced by the method of claim 23.

SEQUENCE LISTING

<110> German, Michael S.
Lin, Joseph

<120> HUMAN NEUROGENIN 3-ENCODING NUCLEOTIDE SEQUENCES

<130> UCSF-129WO

<140> 60/128,180
<141> 1999-04-09

<160> 4

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 5340
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (3022)...(3666)
<223> Coding sequence of human neurogenin3

<400> 1

ggatccctcg tggccagggt tcccttcaag gtgcttagcc aggtcaggag gccctagaga 60
agcatggttt ggattttctt tcccagacca aaaaagctcc aagttggttc tctcccagtt 120
tctaacttgc agttaaataa atcaggcaag gctggcctat gaggcagaca agtgtgaaga 180
aggagaagga ggaggagaag gagaaggaga aagaagaaga aggaggagaa gaagaagaag 240
aagaagaaga agaagaggag gaggaggagg aggaggagga agcagcagca gcagcagcag 300
cttgaatgga cagtggttcc ccttgcctag aaaatgggac cattatttct tttctaatct 360
gacccccaga ctcaggactt cctctatttt ctgcattttg gggtctcttg ttttgccttg 420
aaaaaaaatg ttttctccca aatcaaggag cagtagctgg tgcaagggaa aatctagggc 480
taggagtctt aagatatgac ttctatgtgg ttctgataga acttgctggg tgaccttgag 540
agagtcactc cccctctctg ggccttgatt ttttcatctt taaagaaggc ctcaaattcc 600
cattcttatg agaagaagac aagctcctag tgagtggtga cctaagggag cagctgcagc 660
aaaatgctaa cctgacagtc ccagatggtc cctttattgg ttctgaccct ggtctcaggc 720
ttcatttccc cacagcaagg gaaggagcct gctcacagag caccagctaa gatcagcagg 780
accgcgccac acccccgccc agtcctagag ccccctctc gctggttcct gagcatacca 840
ccctcttcct tggaggaaaa tttgccccca agcagcctag gcggtaagag gctatcacta 900
gggcagactc acagacctac ctcatccccc cacccccacc tacagtctcg aagtcgggtc 960
ctgtcccctc ctgcagtttc cgggagactc aggatatctg gacctgctag aaagagaagc 1020
cttcctcgcc taaggagact taaaccggga tacttaaacc tcccgcctcg gcgtcttcct 1080
ccaggcacga ccgggtcaag agagagaagc ggaagctgca acccctcact ctgagtgacc 1140
ggaagcagaa gaccacggga tgtcccaggc ggggacaaga ggaggggctg gggaagaaag 1200
gagggatgat gagttcagag tccctttgga aaggtttcca gagagcgcta ccagggacaa 1260
cccaaggggc tggggaagtc cctgccttgt gctctctgtg cgatgcccga gtgatgcaga 1320
ggcagggggc tggagcaggt gactgctggc agctgctgtc tgtctgtgat tggaccggag 1380
gactaagggg agaaaaagtt tatcagcttc tcccagtgcc tgcacgctgt ggtagttcaa 1440
aagacacgag ggggaggggc acagcagctc tgcttcccag cgccttggga gactgaagtg 1500
aaaggaacgc ttgagcccag gagttcgaga ccatcctggg caacaaagca agaccgcccc 1560
tcaccccata caaaataaaa atacaaataa attagccggg cacagtggcg catgcctgta 1620
gtctcagcta ctgggaaggc tgaagtggga ggatagcttg agcccaggag atcaaggctg 1680
cagtgagctg tgattgcacc actgcagtcc agcctgggcg acagaaggag accgtttttt 1740
ggttttgttt gttcgtttaa aaaaaaaaag aagcaagagc tcactgtgaa ctcctggttc 1800
cttcctcccc tcctcacact tcccagaact cttcctgtca cggttcctgg ccagaacgct 1860
gggatactat ctacaagctg tagtaggctt gtagtaatgg aatgtccgct tgaggggtcc 1920
ccgcacagcc aaccccggcc tctggagtgg gatctatggg ggtggggttc taagcgcctc 1980
tggggagtgt gaggtagcat ctcagggtgt ggcagaggct cggacacccc caaaaggtct 2040

```
gtgaatggaa gggacatagg caggatctct ctcagtgatg tcccctgtct tccaggatga   2100
agagaggcag tgaaacacca ggagagcagg gcgtccttta gaattcctgg acccttctcc   2160
aggctgctag tcaggacaat gagctcgtgg ttgtctttgc cactatcttc ctgtgcgatt   2220
tcagacaagc cacctccctc actaagccta aatttcccca tgtgtaacgt gcaggcattg   2280
taccctagag gcatcaaagt cccctccagg acagatgcta aggaaagata ggctaggagc   2340
aaagccgtct gaggtggcct gaccagagcc acacgaggct cttctcactg ggcgaggctc   2400
tttgaggaac cgagagttgc tgggacccag cccgccctcg agagagcaaa cagagcggcg   2460
ctcccctccc ccgaccccgg ccctttgtcc ggaatccagc tgtgctgcgg gggaggagcg   2520
ggctcgcgtg gcgcggcccc agggccccgg cgctgattgg ccggtggcgc gggcagcagc   2580
cgggcaggca cgctcctggc ccgggcgaag cagataaagc gtgccaaggg gcacacgact   2640
tgctgctcag gaaatccctg cggtctcacc gccgcgcctc gagagagagc gtgacagagg   2700
cctcggaccc cattctctct tcttttctcc tttggggctg gggcaactcc caggcggggg   2760
cgcctgcagc tcagctgaac ttggcgacca gaagcccgct gagctcccca cggccctcgc   2820
tgctcatcgc tctctattct tttgcgccgg tagaaaggta atatttggag gcctccgagg   2880
gacgggcagg ggaaagaggg atcctctgac ccagcggggg ctgggaggat ggctgttttt   2940
gtttttccc acctagcctc ggaatcgcgg actgcgccgt gacggactca aacttaccct   3000
tccctctgac cccgccgtag g atg acg cct caa ccc tcg ggt gcg ccc act      3051
                        Met Thr Pro Gln Pro Ser Gly Ala Pro Thr
                         1               5                  10

gtc caa gtg acc cgt gag acg gag cgg tcc ttc ccc aga gcc tcg gaa      3099
Val Gln Val Thr Arg Glu Thr Glu Arg Ser Phe Pro Arg Ala Ser Glu
                15                  20                  25

gac gaa gtg acc tgc ccc acg tcc gcc ccg ccc agc ccc act cgc aca      3147
Asp Glu Val Thr Cys Pro Thr Ser Ala Pro Pro Ser Pro Thr Arg Thr
            30                  35                  40

cgg ggg aac tgc gca gag gcg gaa gag gga ggc tgc cga ggg gcc ccg      3195
Arg Gly Asn Cys Ala Glu Ala Glu Glu Gly Gly Cys Arg Gly Ala Pro
        45                  50                  55

agg aag ctc cgg gca cgg cgc ggg gga cgc agc cgg cct aag agc gag      3243
Arg Lys Leu Arg Ala Arg Arg Gly Gly Arg Ser Arg Pro Lys Ser Glu
    60                  65                  70

ttg gca ctg agc aag cag cga cgg agt cgg cga aag aag gcc aac gac      3291
Leu Ala Leu Ser Lys Gln Arg Arg Ser Arg Arg Lys Lys Ala Asn Asp
75                  80                  85                  90

cgc gag cgc aat cga atg cac aac ctc aac tcg gca ctg gac gcc ctg      3339
Arg Glu Arg Asn Arg Met His Asn Leu Asn Ser Ala Leu Asp Ala Leu
                95                  100                 105

cgc ggt gtc ctg ccc acc ttc cca gac gac gcg aag ctc acc aag atc      3387
Arg Gly Val Leu Pro Thr Phe Pro Asp Asp Ala Lys Leu Thr Lys Ile
            110                 115                 120

gag acg ctg cgc ttc gcc cac aac tac atc tgg gcg ctg act caa acg      3435
Glu Thr Leu Arg Phe Ala His Asn Tyr Ile Trp Ala Leu Thr Gln Thr
        125                 130                 135

ctg cgc ata gcg gac cac agc ttg tac gcg ctg gag ccg ccg gcg ccg      3483
Leu Arg Ile Ala Asp His Ser Leu Tyr Ala Leu Glu Pro Pro Ala Pro
    140                 145                 150

cac tgc ggg gag ctg ggc agc cca ggc ggt tcc ccc ggg gac tgg ggg      3531
His Cys Gly Glu Leu Gly Ser Pro Gly Gly Ser Pro Gly Asp Trp Gly
155                 160                 165                 170
```

```
tcc ctc tac tcc cca gtc tcc cag gct ggc agc ctg agt ccc gcc gcg     3579
Ser Leu Tyr Ser Pro Val Ser Gln Ala Gly Ser Leu Ser Pro Ala Ala
                175                 180                 185

tcg ctg gag gag cga ccc ggg ctg ctg ggg gcc acc tct tcc gcc tgc     3627
Ser Leu Glu Glu Arg Pro Gly Leu Leu Gly Ala Thr Ser Ser Ala Cys
            190                 195                 200

ttg agc cca ggc agt ctg gct ttc tca gat ttt ctg tga aaggacctgt      3676
Leu Ser Pro Gly Ser Leu Ala Phe Ser Asp Phe Leu  *
        205                 210

ctgtcgctgg gctgtgggtg ctaagggtaa gggagaggga gggagccggg agccgtagag   3736
ggtggccgac ggcggcggcc ctcaaaagca cttgttcctt ctgcttctcc ctggctgacc   3796
cctggccggc ccaggctcca cgggggcggc aggctgggtt cattccccgg ccctccgagc   3856
cgcgccaacg cacgcaaccc ttgctgctgc ccgcgcgaag tgggcattgc aaagtgcgct   3916
cattttaggc ctcctctctg ccaccacccc ataatctcat tcaaagaata ctagaatggt   3976
agcactaccc ggccggagcc gcccaccgtc ttgggtcgcc ctaccctcac tcaagtctgt   4036
ctgcctctca gtctcttacc acccctcctc caatgtgatt caatccaatg tttggtctct   4096
cagcgcttac tcccccttgcc ttgctccaaa gacgctgccg atctgctcta ctcccaatca   4156
ggtccgggat ttcagggcgc ctcactctgc cttaaagcca cgaaggcgac cctctgcctt   4216
ctcctcgtgc acttttcgga gccattgccc tcccggggcg gaagaccagg ctgtgaactg   4276
ggaaagcgct agcccggcca gggagcatct ccccagcctc cctgcgaact gcgcctgaaa   4336
cgtgagctgc gctgcaggtg cctggagcac cgcgcatctt ttttttttaa atctgtttgt   4396
aaattatatg atgccttttg aaatcaattt tggtacagta aaattatatg gcccctcccc   4456
tgttttacac atttgtattt attaatgaga tttcacagca gggaaaagcc tatattttgg   4516
atattagatt atttagggat tgctggatga catttaagcc aataaaaaaa aatggacctt   4576
caagaagcct tggcaagatg actccattgt gtgttgggga gaggagggcc acagtcacta   4636
cagctgagga agagcacttc tgtccaaaga gagggatgac actctttctg gaggtctggg   4696
ctagagccag ggcagattgg gtttggagag ctggaagtct tctaagtaat tattggtcca   4756
gctccctttt ttctatatag ggcaatgact cctcttattt caaagagtgg tttagaagaa   4816
agacaagcct ccaactagga caactgactc tcacttgctg gccctttccc caactccacc   4876
agcctagctt tagagcaact gttggttgca cttggggaag ggatacagta ataattcaat   4936
tgcagagtca gagtcctcgg aaacacggct gggctgggca tcctaggaat tttcccaagg   4996
tgcttagagg cctagcaaat cccctgagca tattttactc cccaggcact gaggtggctg   5056
tgtcgtgaac tccttgaact gagcagccag gagcaaagaa ggtggagcgt ctggctggaa   5116
tatccagcaa cgccccctcc ctcatcacct ggcagccttg attgaaaact tattaagaaa   5176
ctgttcaagg tttccagcca caccatgtct cttactggca aggtggaata ggactggtgc   5236
agcatgagca ctgaaatctg tcccaggagt gccagtagag caccactaca tgacttcagg   5296
gacccctagg acctcagaga atatggtcta agctgtaagg atcc                    5340
```

<210> 2
<211> 214
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Thr Pro Gln Pro Ser Gly Ala Pro Thr Val Gln Val Thr Arg Glu
 1               5                  10                  15
Thr Glu Arg Ser Phe Pro Arg Ala Ser Glu Asp Glu Val Thr Cys Pro
            20                  25                  30
Thr Ser Ala Pro Pro Ser Pro Thr Arg Thr Arg Gly Asn Cys Ala Glu
        35                  40                  45
Ala Glu Glu Gly Gly Cys Arg Gly Ala Pro Arg Lys Leu Arg Ala Arg
    50                  55                  60
Arg Gly Gly Arg Ser Arg Pro Lys Ser Glu Leu Ala Leu Ser Lys Gln
65                  70                  75                  80
Arg Arg Ser Arg Arg Lys Lys Ala Asn Asp Arg Glu Arg Asn Arg Met
                85                  90                  95
His Asn Leu Asn Ser Ala Leu Asp Ala Leu Arg Gly Val Leu Pro Thr
            100                 105                 110
```

```
Phe Pro Asp Asp Ala Lys Leu Thr Lys Ile Glu Thr Leu Arg Phe Ala
        115                 120                 125
His Asn Tyr Ile Trp Ala Leu Thr Gln Thr Leu Arg Ile Ala Asp His
    130                 135                 140
Ser Leu Tyr Ala Leu Glu Pro Pro Ala Pro His Cys Gly Glu Leu Gly
145                 150                 155                 160
Ser Pro Gly Gly Ser Pro Gly Asp Trp Gly Ser Leu Tyr Ser Pro Val
                165                 170                 175
Ser Gln Ala Gly Ser Leu Ser Pro Ala Ala Ser Leu Glu Glu Arg Pro
            180                 185                 190
Gly Leu Leu Gly Ala Thr Ser Ser Ala Cys Leu Ser Pro Gly Ser Leu
        195                 200                 205
Ala Phe Ser Asp Phe Leu
    210
```

<210> 3
<211> 1861
<212> DNA
<213> M. musculus

<220>
<221> CDS
<222> (1093)...(1737)

<400> 3

```
ggatcccaag gtgatattga acctggccaa gcaatagttt ctgagtagaa aggacttgag     60
cagggaccgt ctctggtcac tctgtcctct ttcccaggat ggagtcagtc tgtgaaacat    120
ggttgcacac acatttcctg acccaaccca tagtggcgga gagctggata gcactttgaa    180
ctaatgggcg ctcctcccag ctgccagcca agaagacact tgactccttg atcgctggtt    240
catttagaca agccgtttcc ctctctgagc caaaagaccc catgtgtaat actcaaagaa    300
gaggccttcc ttatatatat ataggcaccc ccaaacctcc ttcatgctac caagaaaggg    360
tctggacaca tgccaaaaag aaagaggaaa aggcaaagct ctccccagcg gccggacggg    420
actcttctgg ctgggcgagg ctctttgagg aaccgagagt tgctgggact gagcccgcga    480
cgggggaggc gtggagtggg ggaacaaaca gagtgctgct cccctccccc gacccctgcc    540
ctttgtccgg aatccagctg tgctctgcgg gtgggggttg tgggggagg agcgggctcg    600
cgtggcgcag cccctgggcc ccctccgctg attggcccgt ggtgcaggca gcagcccggc    660
aggcacgctc ctggccgggg gcagagcaga taaagcgtgc caggggacac acgacttgca    720
tgcagctcag aaatccctct gggtctcatc actgcagcag tggtcgagta cctcctcgga    780
gcttttctac gacttccaga cgcaatttac tccaggcgag ggcgcctgca gtttagcaga    840
acttcagagg gagcagagag gctcagctat ccactgctgc ttgacactga ccctatccac    900
tgctgcttgt cactgactga cctgctgctc tctattcttt tgagtcggga gaactaggta    960
acaattcgga aactccaaag ggtggatgag gggcgcgcgg ggtgtgtgtg ggggatactc   1020
tggtcccccg tgcagtgacc tctaagtcag aggctggcac acacacacct tccatttttt   1080
cccaaccgca gg atg gcg cct cat ccc ttg gat gcg ctc acc atc caa gtg   1131
              Met Ala Pro His Pro Leu Asp Ala Leu Thr Ile Gln Val
               1               5                  10

tcc cca gag aca caa caa cct ttt ccc gga gcc tcg gac cac gaa gtg    1179
Ser Pro Glu Thr Gln Gln Pro Phe Pro Gly Ala Ser Asp His Glu Val
     15                  20                  25

ctc agt tcc aat tcc acc cca cct agc ccc act ctc ata cct agg gac    1227
Leu Ser Ser Asn Ser Thr Pro Pro Ser Pro Thr Leu Ile Pro Arg Asp
 30                  35                  40                  45

tgc tcc gaa gca gaa gtg ggt gac tgc cga ggg acc tcg agg aag ctc    1275
Cys Ser Glu Ala Glu Val Gly Asp Cys Arg Gly Thr Ser Arg Lys Leu
                 50                  55                  60

cgc gcc cga cgc gga ggg cgc aac agg ccc aag agc gag ttg gca ctc    1323
Arg Ala Arg Arg Gly Gly Arg Asn Arg Pro Lys Ser Glu Leu Ala Leu
             65                  70                  75
```

```
agc aaa cag cga aga agc cgg cgc aag aag gcc aat gat cgg gag cgc      1371
Ser Lys Gln Arg Arg Ser Arg Arg Lys Lys Ala Asn Asp Arg Glu Arg
         80                  85                  90

aat cgc atg cac aac ctc aac tcg gcg ctg gat gcg ctg cgc ggt gtc      1419
Asn Arg Met His Asn Leu Asn Ser Ala Leu Asp Ala Leu Arg Gly Val
     95                 100                 105

ctg ccc acc ttc ccg gat gac gcc aaa ctt aca aag atc gag acc ctg      1467
Leu Pro Thr Phe Pro Asp Asp Ala Lys Leu Thr Lys Ile Glu Thr Leu
110                 115                 120                 125

cgc ttc gcc cac aac tac atc tgg gca ctg act cag acg ctg cgc ata      1515
Arg Phe Ala His Asn Tyr Ile Trp Ala Leu Thr Gln Thr Leu Arg Ile
                130                 135                 140

gcg gac cac agc ttc tat ggc ccg gag ccc cct gtg ccc tgt gga gag      1563
Ala Asp His Ser Phe Tyr Gly Pro Glu Pro Pro Val Pro Cys Gly Glu
            145                 150                 155

ctg ggg agc ccc gga ggt ggc tcc aac ggg gac tgg ggc tct atc tac      1611
Leu Gly Ser Pro Gly Gly Gly Ser Asn Gly Asp Trp Gly Ser Ile Tyr
        160                 165                 170

tcc cca gtc tcc caa gcg ggt aac ctg agc ccc acg gcc tca ttg gag      1659
Ser Pro Val Ser Gln Ala Gly Asn Leu Ser Pro Thr Ala Ser Leu Glu
    175                 180                 185

gaa ttc cct ggc ctg cag gtg ccc agc tcc cca tcc tat ctg ctc ccg      1707
Glu Phe Pro Gly Leu Gln Val Pro Ser Ser Pro Ser Tyr Leu Leu Pro
190                 195                 200                 205

gga gca ctg gtg ttc tca gac ttc ttg tga agagacctgt ctggctctgg       1757
Gly Ala Leu Val Phe Ser Asp Phe Leu  *
                210

gtggtgggtg ctagtggaaa gggaggggac cagagccgtc tggagtggga ggtagtggag   1817
gctctcaagc atctcgcctc ttctggcttt cactacttgg atcc                    1861
```

<210> 4
<211> 214
<212> PRT
<213> M. musculus

<400> 4

```
Met Ala Pro His Pro Leu Asp Ala Leu Thr Ile Gln Val Ser Pro Glu
 1               5                  10                  15
Thr Gln Gln Pro Phe Pro Gly Ala Ser Asp His Glu Val Leu Ser Ser
            20                  25                  30
Asn Ser Thr Pro Pro Ser Pro Thr Leu Ile Pro Arg Asp Cys Ser Glu
        35                  40                  45
Ala Glu Val Gly Asp Cys Arg Gly Thr Ser Arg Lys Leu Arg Ala Arg
    50                  55                  60
Arg Gly Gly Arg Asn Arg Pro Lys Ser Glu Leu Ala Leu Ser Lys Gln
65                  70                  75                  80
Arg Arg Ser Arg Arg Lys Lys Ala Asn Asp Arg Glu Arg Asn Arg Met
                85                  90                  95
His Asn Leu Asn Ser Ala Leu Asp Ala Leu Arg Gly Val Leu Pro Thr
            100                 105                 110
Phe Pro Asp Asp Ala Lys Leu Thr Lys Ile Glu Thr Leu Arg Phe Ala
        115                 120                 125
```

```
His Asn Tyr Ile Trp Ala Leu Thr Gln Thr Leu Arg Ile Ala Asp His
    130                 135                 140
Ser Phe Tyr Gly Pro Glu Pro Pro Val Pro Cys Gly Glu Leu Gly Ser
145                 150                 155                 160
Pro Gly Gly Gly Ser Asn Gly Asp Trp Gly Ser Ile Tyr Ser Pro Val
                165                 170                 175
Ser Gln Ala Gly Asn Leu Ser Pro Thr Ala Ser Leu Glu Glu Phe Pro
            180                 185                 190
Gly Leu Gln Val Pro Ser Ser Pro Ser Tyr Leu Leu Pro Gly Ala Leu
        195                 200                 205
Val Phe Ser Asp Phe Leu
    210
```

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/08436

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) :C07K 14/435
US CL :530/350
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MEDLINE, BIOSIS, USPATFUL
search terms: neurogenin#, Ngn3

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- A	US 5,795,723 A (TAPSCOTT et al) 18 AUGUST 1998, cols. 4, 28 & 30.	1 --- 2-3

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
05 JULY 2000	08 AUG 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer ROBERT C. HAYES, PH.D. Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/08436

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-3

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/08436

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-3, drawn to isolated human neurogenin-3 polypeptides.
Group II, claim(s) 4-10 and 14-18, drawn to polynucleotides encoding a human neurogenin-3 polypeptide, expression vectors, host cells, and methods of producing these polypeptides.
Group III, claim(s) 11, drawn to antibodies of a human neurogenin-3 polypeptide.
Group IV, claim(s) 12, drawn to methods of detecting homologous human neurogenin-3 polynucleotides.
Group V, claim(s) 13, drawn to methods of detecting an islet cell precursor comprising the detection of neurogenin-3 expression.
Group VI, claim(s) 19, drawn to a method of identifying an agent that modulates human neurogenin-3 activity.
Group VII, claim(s) 20-22, drawn to methods of detecting a predisposition to pancreatic islet cell defects that are associated with altered neurogenin-3 activity/genomic sequences.
Group VIII, claim(s) 23, drawn to a method of producing a human pancreatic islet cell.
Group IX, claim(s) 24, drawn to a pancreatic islet cell.

The inventions listed as Groups I-IX do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Group I is directed to a human neurogenin-3 polypeptide, which is the first product. However, because Tapscott et al teach a human neurogenin-3/neuroD3 polypeptide with sufficient homology to the human neurogenin-3 polypeptide of SEQ ID NO: 2 to be considered a human neurogenin-3 polypeptide "variant" or "analog", no special technical feature exists for Group I as defined by PCT Rule 13.2, because it does not define a contribution over the prior art. Groups I-III and IX are also drawn to a structurally different products, which do not require each other for their practice and do not share the same or a corresponding technical feature. The technical features of Groups IV-VIII are drawn to methods having different goals, method steps and starting materials, which do not require each other for their practice and do not share the same or a corresponding technical feature. Note that PCT Rule 13 does not provide for multiple products or methods within a single application. Because the technical feature of Group I is not a special technical feature, and because the technical features of the Group II-IX inventions are not present in the Group I claims, unity of invention is lacking.

US005695995A

United States Patent [19]

Weintraub et al.

[11] Patent Number: 5,695,995

[45] Date of Patent: Dec. 9, 1997

[54] NEUROGENIC DIFFERENTIATION (NEUROD) GENES

[75] Inventors: **Harold M. Weintraub**, deceased, late of Seattle, Wash., by Nancy Weintraub, executrix; **Jacqueline E. Lee**, Denver, Colo.; **Stanley M. Hollenberg**, Portland, Oreg.; **Stephen J. Tapscott**, Seattle, Wash.

[73] Assignee: **Fred Hutchinson Cancer Research Center**, Seattle, Wash.

[21] Appl. No.: **552,142**

[22] Filed: **Nov. 2, 1995**

Related U.S. Application Data

[63] Continuation-in-part of PCT/US95/05741, May 8, 1995, which is a continuation-in-part of Ser. No. 239,238, May 6, 1994, abandoned.

[51] **Int. Cl.**[6] **C12N 15/85**; C12N 5/10; C12N 15/18

[52] **U.S. Cl.** **435/325**; 435/69.1; 435/69.4; 435/172.3; 435/252.33; 435/320.1; 435/357; 435/360; 536/23.1; 536/23.5; 536/23.51

[58] **Field of Search** 435/69.1, 69.4, 435/172.3, 320.1, 240.2, 357, 325, 360, 252.33; 536/23.1, 23.5, 23.51

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,302,549	4/1994	Blackwood et al. .
5,322,801	6/1994	Kingston et al. .
5,352,595	10/1994	Tapscott et al. .

OTHER PUBLICATIONS

Wilson, P.A. et al., "Induction of epidermis and inhibition of neural fate by Bmp–4," *Nature*, 376:331–333 (1995).

Lee, J.E. et al., "Conversion of *Xenopus* Ectoderm into Neurons by NeuroD, a basic Helix–Loop–Helix Protein," *Science*, 268:836–844 (1995).

Shimizu, C. et al., "MATH–2, a mammalian helix–loop–helix factor structurally related to the product of *Drosophila* proneural gene *atonal*, is specifically expressed in the nervous sysstem," *Eur. J. Biochem.*, 229:239–248 (1995).

Naya, F. J. et al., "Tissue–specific regulation of the insulin gene by a nove basic helix–loop–helix transcription factor," *Genes and Development*, 9:1009–1019 (1995).

Singson, A. et al., "Direct downstream targets of proneural activators in the imaginal disc include genes involved in lateral inhibitory signaling," *Genes & Development*, 8:2058–2071 (1994).

Turner, D.L. et al., "Expression of achaete–scute homolog 3 in *Xenopus* embryos converts ectodermal cells to a neural fate," *Genes & Development*, 8:1434–1447 (1994).

Rupp, R.A.W. et al., "*Xenopus embryos regulate the nuclear localization of XMyoD*," *Genes & Development*, 8:1311–1323 (1994).

Buffinger, N. et al., "Myogenic specification in somites: induction by axial structures," *Development*, 120:1443–1452 (1994).

Kramatschek, B. et al., "Neuroectodermal transcription of the *Drosophila* neurogenic genes E(spl) and HLH–m5 is regulated by proneural genes," *Development*, 120:815–826 (1994).

Burcin, M. et al., "Factors influencing nuclear receptors in transcriptional repression," *Cancer Biology*, 5:337–346 (1994).

Rudnicki, M.A., et al., "MyoD or Myf–5 Is Required for the Formation of Skeletal Muscle," *Cell*, 75:1351–1359 (1993).

Guillemot, F. et al., "Mammalian achaete–scute Homolog 1 is Required for the Early Development of Olfactory and Autonomic Neurons," *Cell*, 75:463–476 (1993).

Lamb, T.M. et al., "Neural Induction by the Secreted Polypeptide Noggin," *Science*, 262:713–718 (1993).

Hasty, P. et al., "Muscle deficiency and neonatal death in mice with a targeted mutation in the *myogenin* gene," *Nature*, 364:501–506 (1993).

Nabeshima, Y. et al., "*Myogenin* gene disruption results in perinatal lethality because of severe muscle defect," *Nature*, 364:532–535 (1993).

Guillemot, F. et al., "Dynamic expression of the murine Achaete–Scute homologue Mash–1 in the developing nervous system," *Mechanisms of Development*, 42:171–185 (1993).

Rong, P.M. et al., "The neural tube/notochord complex is necessary for vertebral but not limb and body wall striated muscle differentiation," *Development*, 115:657–672 (1992).

Lo, L.C. et al., "Mammalian achaete–scute homolog 1 is transiently expressed by spatially restricted subsets of early neuroepithelial and neural crest cells," *Genes & Development*, 5:1524–1537 (1991).

Blackwood, E.M. et al., "Max: A Helix–Loop–Helix Zipper Protein That Forms a Sequence–Specific DNA–Binding Complex with Myc," *Science*, 251:1211–1217 (1991).

Johnson, J.E. et al., "Two rat homologues of *Drosophila achaete–scute* specifically expressed in neuronal precursors," *Nature*, 346:858–861 (1990).

Maniatis et al., "Molecular cloning: a laboratory manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, p. 227.

Gen Bank NCBI gibbsq 109439, 1992, cited as MPSEARCH Result 11.

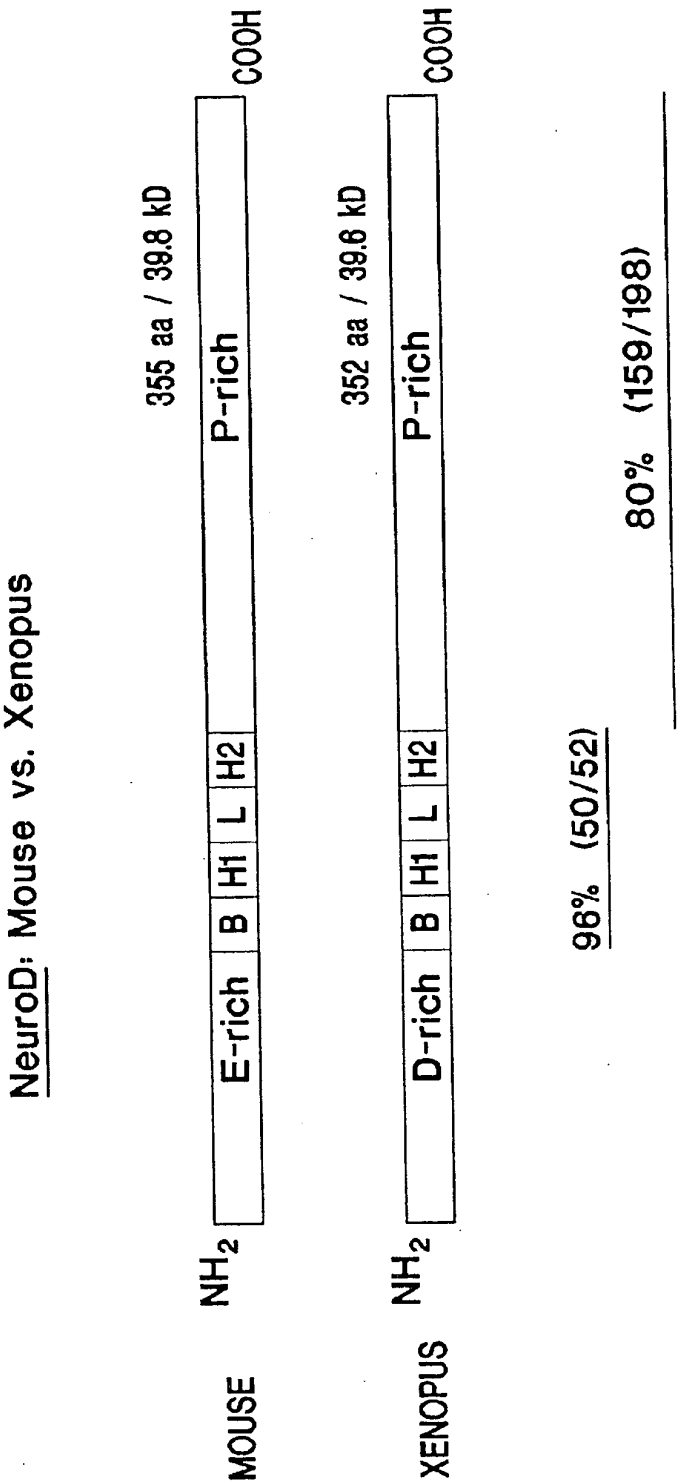
Primary Examiner—John L. LeGuyader

Attorney, Agent, or Firm—Christensen O'Connor Johnson & Kindness PLLC

[57] ABSTRACT

Neurogenic differentiation genes and proteins are identified, isolated, and sequenced. Expression of neuroD has been demonstrated in neural, pancreatic, and gastrointestinal cells. Ectopic expression of neuroD in non-neuronal cells of Xenopus embryos induced formation of neurons.

8 Claims, 1 Drawing Sheet


NeuroD: Mouse vs. Xenopus
355 aa / 39.8 kD
352 aa / 39.8 kD
COOH
COOH
P-rich
P-rich
E-rich
D-rich
B
H1
L
H2
NH2
NH2
MOUSE
XENOPUS
98% (50/52)
80% (159/198)

NEUROGENIC DIFFERENTIATION (NEUROD) GENES

This application is a continuation-in-part of international application No. PCT/US95/05741, filed May 8, 1995, which is a continuation-in-part of parent application U.S. Ser. No. 08/239,238, filed May 6, 1994 (abandoned).

This invention was made with government support under grant CA42506 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD OF THE INVENTION

The invention relates to molecular biology and in particular to genes and proteins involved in vertebrate neural development.

BACKGROUND OF THE INVENTION

There are currently several examples of transcription regulatory proteins sharing a basic helix-loop-helix (bHLH) secondary structure. bHLH proteins form homodimeric and homodimeric complexes binding DNA in the 5' regulatory regions of genes controlling expression. Among the bHLH proteins, mammalian MyoD and Drosophila AS-C are presently thought to play developmental roles in muscle development and in sensory organ development, respectively. Both proteins are thought to exert their effects by binding 5' regulatory nucleotide sequences in genes that seem specifically determinative of cellular differentiation and fate. However, the specific developmental roles of the genes affected by MyoD and AS-C remain largely unknown, as are the molecular details of the developmental pathways regulated by these genes.

The presently disclosed NeuroD proteins represent a new sub-family of bHLH proteins and are implicated in vertebrate neuronal, endocrine and gastrointestinal development.

Subsequent to the filing of the parent application, Naya et al. (*Genes & Devel.* 9: 1009–1019, 1995) disclosed the isolation of a hamster bHLH transcription factor (BETA2) that binds to the insulin E-box sequence. Shimizu et al. (*Eur. J. Biochem.* 229: 239–248, 1995) disclosed the isolation of a mouse HLH protein, MATH-2 that is detected in neural tissue. Comparison of these sequences with the neuroD sequences disclosed herein demonstrate that they are members of the NeuroD family of proteins.

Neural tissues and endocrine tissues do not regenerate. Damage is permanent. Paralysis, loss of vision or hearing and hormonal insufficiency are also permanent. Tumors in neural and endocrine tissues can also be very difficult to treat because of the toxic side effects that conventional chemotherapeutic drugs may have on nervous tissues. The medical community and public would greatly benefit from the availability of agents active in triggering differentiation in neuroectodermal stem cells. Such neuronal differentiating agents could be used for construction of test cell lines, assays for identifying candidate therapeutic agents capable of inducing regeneration of neuronal and endocrine tissues, gene therapy, and differentiation of tumor cells.

SUMMARY OF THE INVENTION

Mammalian and amphibian NeuroD proteins were identified, and polynucleotide molecules encoding NeuroD proteins were isolated and sequenced. neuroD genes encode proteins that are distinctive members of the bHLH family. In addition, the present invention provides a family of NeuroD proteins that share a highly conserved HLH region. Representative polynucleotide molecules encoding members of the NeuroD family include neuroD, neuroD2 and neuroD3.

A representative nucleotide sequence encoding murine neuroD is shown in SEQ ID NO:1. The HLH coding domain of murine neuroD resides between nucleotides 577 and 696 in SEQ ID NO:1. The deduced amino acid sequence of murine NeuroD is shown in SEQ ID NO:2. There is a highly conserved region following the helix-2 domain from amino acid 150 through amino acid 199 of SEQ ID NO:2 that is not shared by other bHLH proteins.

A representative nucleotide sequence encoding Xenopus neuroD is shown in SEQ ID NO:3. The HLH coding domain of Xenopus neuroD resides between nucleotides 376 and 495 in SEQ ID NO:3. The deduced amino acid sequence of Xenopus NeuroD is shown in SEQ ID NO:4. There is a highly conserved region following the helix-2 domain from amine acid 157 through amine acid 199 of SEQ ID NO:4 that is not shared by other bHLH proteins.

Representative nucleotide and deduced amine acid sequences of the human NeuroD family are shown in SEQ ID NOS:8–15. Representative nucleotide and deduced amine acid sequences of a human homolog of murine neuroD are shown in SEQ ID NOS:8 and 9 (partial genomic sequence) and SEQ ID NOS:14 and 15 (human cDNA). Representative nucleotide and deduced amine acid sequences of the human and murine neuroD2 are shown in SEQ ID NOS:10 and 11, and 16 and 17, respectively. Representative nucleotide and deduced amine acid sequences for human neuroD3 are shown in SEQ ID NOS:12 and 13. The disclosed human clones, 9F1(and its corresponding cDNA HC2A; now referred to as human neuroD) and 14B1(now referred to as human neuroD2), have an identical HLH motif: amine acid residues 117–156 in SEQ ID NO:9 and 15, and residues 137–176 in SEQ ID NO:11 (corresponding to nucleotides 405–524 of SEQ ID NO:8 and SEQ ID NO:14, and nucleotides 463–582 of SEQ ID NO:10). Comparison of the deduced amine acid sequences of these neuroD genes shows that human NeuroD3 contains an HLH domain between amine acid residues 50–89 of SEQ ID NO:13 (corresponding to nucleotides 149–268 of SEQ ID NO:12) and that murine NeuroD2 contains an HLH domain between amine acids residues 138–177 of SEQ ID NO:17 (corresponding to nucleotides 642–761 of SEQ ID NO:16). The HLH domain of murine NeuroD2 is identical to the human NeuroD and human NeuroD2.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 schematically depicts the domain structure of the murine and Xenopus NeuroD bHLH proteins.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Tissue-specific bHLH proteins that regulate early neuroectodermal differentiation were discovered using expression cloning and screening assays designed to identify possible bHLH proteins capable of interacting with the protein product of the *Drosophila daughterless* gene. These proteins belong to a family of proteins that share conserved residues in the HLH region. The term NeuroD is generally used to encompass all members of the NeuroD family. and includes neuroD, neuroD2 and neuroD3 coding sequences and proteins.

NeuroD proteins are transiently expressed in differentiating neurons during embryogenesis. NeuroD is also detected in adult brain, in the granule layer of the hippocampus and

the cerebellum. In addition, murine neuroD expression has been detected in the pancreas and gastrointestinal tissues of developing embryos and post-natal mice. NeuroD contains the basic helix-loop-helix (bHLH) domain structure that has been implicated in the binding of bHLH proteins to upstream recognition sequences and activation of downstream target genes. The present invention provides representative NeuroD proteins, which include the murine NeuroD protein of SEQ ID NO:2, the amphibian NeuroD protein of SEQ ID NO:4, murine NeuroD2 protein of SEQ ID NO:17, human NeuroD protein of SEQ ID NOS:9 and 15, human NeuroD2 protein of SEQ ID NO:11, and human NeuroD3 protein of SEQ D NO:13. Based on homology with other bHLH proteins, the bHLH domain for murine NeuroD is predicted to reside between amino acids 102 and 155 of SEQ ID NO:2, and between amino acids 101 and 157 of SEQ ID NO:4 for the amphibian NeuroD.

As detailed below, the present invention provides the identification of human neuroD and, in addition, provides unexpected homologous genes of the same family based on highly conserved sequences across the HLH domain shared between the two human genes at the amino acid level (neuroD2 and neuroD3; SEQ ID NOS:10 and 11, and 12 and 13, respectively).

NeuroD proteins are transcriptional activators that control transcription of downstream target genes that cause neuronal progenitors to differentiate into mature neurons. In the neurula stage of the mouse embryo (el0), murine neuroD is highly expressed in the neurogenic derivatives of neural crest cells, the cranial and dorsal root ganglia, and postmitotic cells in the central nervous system (CNS). During mouse development, neuroD is expressed transiently and concomitant with neuronal differentiation in differentiating neurons in sensory organs such as in nasal epithelium and retina. In Xenopus embryos ectopic expression of neuroD in non-neuronal cells induced formation of neurons. As discussed in more detail below, NeuroD proteins are expressed in differentiating neurons and are capable of causing the conversion of non-neuronal cells into neurons. The present invention encompasses NeuroD variants that, for example, are modified in a manner that results in a NeuroD protein capable of binding to its recognition site, but unable to activate downstream genes. The present invention also encompasses fragments of NeuroD that, for example, are capable of binding the natural NeuroD partner, but are incapable of activating downstream genes. NeuroD proteins encompass proteins retrieved from naturally occurring materials and closely related, functionally similar proteins retrieved by antisera specific to NeuroD, and recombinantly expressed proteins encoded by genetic materials (DNA, RNA, cDNA) retrieved on the basis of their similarity to the unique regions in the neuroD family of genes.

The present invention provides representative isolated and purified polynucleotide molecules encoding proteins of the NeuroD family. Representative polynucleotide molecules encoding NeuroD include the sequences presented in SEQ ID NOS:1, 3, 8, 10, 12, 14, and 16. Polynucleotide molecules encoding NeuroD include those sequences resulting in minor genetic polymorphisms, differences between species, and those that contain amino acid substitutions, additions, and/or deletions. According to the present invention, polynucleotide molecules encoding NeuroD encompass those molecules that encode NeuroD proteins or peptides that share identity with the sequences shown in SEQ ID NOS:2, 4, 9, 11, 13, 15, and 17. Such molecules will generally share greater than 35% identity at the amino acid level with the disclosed sequences. The polynucleotide molecules of the present invention may share greater identity at the amino acid level across highly conserved regions such as the HLH domain. For example, the deduced amino acid sequences of murine and Xenopus neuroD genes are 96% identical.

In some instances, one may employ such changes in the sequence of a recombinant neuroD to substantially decrease or even increase the biological activity of NeuroD relative to the wild-type NeuroD activity, depending on the intended use of the preparation. Such changes may also be directed towards endogenous neuroD sequences using, for example, gene therapy methods to alter the gene product.

The NeuroD proteins of the present invention are capable of inducing the expression of neuronal-specific genes, such as N-CAM, β-tubulin, and Xen-1, neurofilament M (NF-M), Xen-2, tanabin-1, shaker-1, and frog HSCL, in a frog embryo. As described below, NeuroD activity may be detected when NeuroD is ectopically expressed in frog oocytes following, for example, injection of Xenopus neuroD RNA into one of the two cells in a two-cell stage Xenopus embryo, and monitoring expression of neuronal-specific genes in the injected as compared to un-injected side of the embryo by immunochemistry or in situ hybridization.

"Over-expression" means an increased level of NeuroD protein or neuroD transcripts in a recombinant transformed host cell relative to the level of protein or transcripts in the parental cell from which the host cell is derived.

As noted above, the present invention provides isolated and purified polynucleotide molecules encoding NeuroD and other members of the NeuroD family. The disclosed sequences may be used to identify and isolate neuroD polynucleotide molecules from suitable host cells such as canine, ovine, bovine, caprine, lagomorph, or avian. In particular, the nucleotide sequences encoding the HLH region may be used to identify polynucleotide molecules encoding other proteins of the NeuroD family. Complementary DNA molecules encoding NeuroD family members may be obtained by constructing a cDNA library mRNA from, for example, fetal brain, newborn brain, adult brain and brain tissues. DNA molecules encoding NeuroD family members may be isolated from such a library using the disclosed sequences in standard hybridization techniques (e.g., Sambrook et al., ibid., and Bothwell, Yancopoulos and Alt, ibid.) or by amplification of sequences using polymerase chain reaction (PCR) amplification (e.g. Loh et al., *Science* 243: 217–222, 1989; Frohman et al., *Proc. Natl. Acad Sci. USA* 85: 8998–9002, 1988; and Erlich (ed.), *PCR Technology: Principles and Applications for DNA Amplification*, Stockton Press, 1989; which are incorporated by reference herein in their entirety). In a similar manner, genomic DNA encoding NeuroD may be obtained using probes designed from the sequences disclosed herein. Suitable probes for use in identifying neuroD sequences may be obtained from neuroD-specific sequences that are highly conserved regions between mammalian and amphibian neuroD coding sequences. Primers, for example, from the region encoding the approximately 40 residues following the helix-2 domain are suitable for use in designing PCR primers. Alternatively, oligonucleotides containing specific DNA sequences from a human neuroD coding region may be used within the described methods to identify related human neuroD genomic and cDNA clones. Upstream regulatory regions of neuroD may be obtained using the same methods. Suitable PCR primers are between 7–50 nucleotides in length, more preferably between 15 and 25 nucleotides in length. Alternatively, neuroD polynucleotide molecules may be isolated using standard hybridization techniques with probes

of at least about 7 nucleotides in length and up to and including the full coding sequence. Southern analysis of mouse genomic DNA probed with the murine neuroD cDNA under stringent conditions showed the presence of only one gene, suggesting that under stringent conditions bHLH genes from other protein families will not be identified. Other members of the neuroD family can be identified using degenerate oligonucleotides based on the sequences disclosed herein for PCR amplification or by hybridization at moderate stringency.

The regulatory regions of neuroD may be useful as tissue-specific promoters. Such regulatory regions may find use in, for example, gene therapy to drive the tissue-specific expression of heterologous genes in pancreatic, gastrointestinal, or neural cells, tissues or cell lines. As shown in Example 14, murine neuroD promoter sequences reside within the 1.4 kb 5' untranslated region. Regulatory sequences within this region are identified by comparison to other promoter sequences and/or deletion analysis of the region itself.

A DNA molecule ceding a NeuroD protein is inserted into a suitable expression vector, which is in turn used to transfect or transform a suitable host cell. Suitable expression vectors for use in carrying out the present invention include a promoter capable of directing the transcription of a polynucleotide molecule of interest in a host cell. Representative expression vectors may include both plasmid and/or viral vector sequences. Suitable vectors include retroviral vectors, vaccinia viral vectors, CMV viral vectors, BLUESCRIPT™ vectors, baculovirus vectors, and the like. Promoters capable of directing the transcription of a cloned gene or cDNA may be inducible or constitutive promoters and include viral and cellular promoters. For expression in mammalian host cells, suitable viral promoters include the immediate early cytomegalovirus promoter (Boshart et al., *Cell* 41: 521–530, 1985) and the SV40 promoter (Subramani et al., *Mol. Cell. Biol.* 1: 854–864, 1981). Suitable cellular promoters for expression of proteins in mammalian host cells include the mouse metallothionien-1 promoter (Palmiter et al., U.S. Pat. No. 4,579,821), a mouse V_K promoter (Bergman et al., *Proc. Natl. Acad Sci. USA* 81: 7041–7045, 1983; Grant et al. *Nucleic Acid Res.* 15: 5496, 1987), and tetracycline-responsive promoter (Gossen and Bujard, *Proc. Natl. Acad Sci. USA* 89: 5547–5551, 1992, and Pescini et al., *Biochem. Biophys. Res. Comm.* 202: 1664–1667, 1994). Also contained in the expression vectors, typically, is a transcription termination signal located downstream of the coding sequence of interest. Suitable transcription termination signals include the early or late polyadenylation signals from SV40 (Kaufman and Sharp, *Mol. Cell. Biol.* 2: 1304–1319, 1982), the polyadenylation signal from the Adenovirus 5 elB region, and the human growth hormone gene terminator (DeNoto et al., *Nucleic Acid Res.* 9: 3719–3730, 1981). Mammalian cells, for example, may be transfected by a number of methods including calcium phosphate precipitation Wigler et al., *Cell* 14: 725, 1978; Corsaro and Pearson, *Somatic Cell Genetics* 7: 603, 1981; Graham and Van der Eb, *Virology* 52: 456, 1973), lipofection, microinjection, and electropotation (Neumann et al., *EMBO J.* 1: 8410845, 1982). Mammalian cells can be transduced with viruses such as SV40, CMV, and the like. In the case of viral vectors, cloned DNA molecules may be introduced by infection of susceptible cells with vital particles. Retroviral vectors may be preferred for use in expressing NeuroD proteins in mammalian cells particularly if NeuroD is used for gene therapy (for review, see, Miller et al. *Methods in Enzymology* 217: 581–599, 1994; which is incorporated herein by reference in its entirety). It may be preferable to use a selectable marker to identify cells that contain the cloned DNA. Selectable markers are generally introduced into the cells along with the cloned DNA molecules and include genes that confer resistance to drugs, such as neomycin, hygromycin, and methotrexate. Selectable markers may also complement auxotrophies in the host cell. Yet other selectable markers provide detectable signals, such as beta-galactosidase to identify cells containing the cloned DNA molecules. Selectable markers may be amplifiable. Such amplifiable selectable markers may be used to amplify the number of sequences integrated into the host genome.

As would be evident to one of ordinary skill in the art, the polynucleotide molecules of the present invention may be expressed in *Saccharomyces cerevisiae*, filamentous fungi, and *E coli*. Methods for expressing cloned genes in *Saccharomyces cerevisiae* are generally known in the art (see, "Gene Expression Technology," *Methods in Enzymology*, Vol. 185, Goeddel (ed.), Academic Press, San Diego, Calif., 1990; and "Guide to Yeast Genetics and Molecular Biology." *Methods in Enzymology*, Guthrie and Fink (eds.), Academic Press, San Diego, Calif., 1991; which are incorporated herein by reference). Filamentous fungi may also be used to express the proteins of the present invention; for example, strains of the fungi Aspergillus (McKnight et al., U.S. Pat. No. 4,935,349, which is incorporated herein by reference). Methods for expressing genes and cDNAs in cultured mammalian cells and in *E coli* are discussed in detail in Sambrook et al. (*Molecular Cloning: A Laboratory Manual, Second Edition*, Cold Spring Harbor, N.Y., 1989; which is incorporated herein by reference). As will be evident to one skilled in the art, one can express the protein of the instant invention in other host cells such as avian, insect, and plant cells using regulatory sequences, vectors and methods well established in the literature.

NeuroD proteins produced according to the present invention may be purified using a number of established methods such as affinity chromatography using anti-NeuroD antibodies coupled to a solid support. Fusion proteins of antigenic tag and NeuroD can be purified using antibodies to the tag. Additional purification may be achieved using conventional purification means such as liquid chromatography, gradient centrifugation, and gel electrophoresis, among others. Methods of protein purification are known in the art (see generally, Scopes, R., *Protein Purification*, Springer-Verlag, N.Y., 1982, which is incorporated herein by reference) and may be applied to the purification of recombinant NeuroD described herein.

The term "capable of hybridizing under stringent conditions" as used herein means that the subject nucleic acid molecules (whether DNA or RNA) anneal to an oligonucleotide of 15 or more contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14 or SEQ ID NO:16.

The choice of hybridization conditions will be evident to one skilled in the art and will generally be guided by the purpose of the hybridization, the type of hybridization (DNA-DNA or DNA-RNA), and the level of desired relatedness between the sequences. Methods for hybridization are well established in the literature. See, for example: Sambrook et al., ibid.; Hames and Higgins, eds, *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington D.C., 1985; Berger and Kimmel, eds, *Methods in Enzymology, Vol. 52, Guide to Molecular Cloning Techniques*, Academic Press Inc., New York, N.Y., 1987; and Bothwell, Yancopoulos and Alt, eds, Methods for *Clon-*

ing and Analysis of Eukaryotic Genes, Jones and Bartlett Publishers, Boston, Mass. 1990; which are incorporated by reference herein in their entirety. One of ordinary skill in the an realizes that the stability of nucleic acid duplexes will decrease with an increased number and location of mismatched bases; thus, the stringency of hybridization may be used to maximize or minimize the stability of such duplexes. Hybridization stringency can be altered by: adjusting the temperature of hybridization; adjusting the percentage of helix-destabilizing agents, such as formamide, in the hybridization mix; and adjusting the temperature and salt concentration of the wash solutions. In general, the stringency of hybridization is adjusted during the post-hybridization washes by varying the salt concentration and/or the temperature. Stringency of hybridization may be reduced by reducing the percentage of formamide in the hybridization solution or by decreasing the temperature of the wash solution. High stringency conditions may involve high temperature hybridization (e.g., 65°–68° C. in aqueous solution containing 4–6× SSC, or 42° C. in 50% formamide) combined with high temperature (e.g., 5°–25° C. below the T_m) and a low salt concentration (e.g., 0.1× SSC). Reduced stringency conditions may involve lower hybridization temperatures (e.g., 35°–42° C. in 20–50% formamide) with intermediate temperature (e.g., 40°–60° C.) and washes in a higher salt concentration (e.g., 2–6× SSC). Moderate stringency conditions, which may involve hybridization at a temperature between 50° C. and 55° C. and washes in 0.1× SSC, 0.1% SDS at between 50° C. and 55° C., may be used to identify clones encoding members of the NeuroD family.

The invention provides isolated and purified polynucleotide molecules encoding NeuroD proteins that are capable of hybridizing under stringent conditions to an oligonucleotide of 15 or more contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, and/or SEQ ID NO:16, including theft complementary strands. The subject isolated neuroD polynucleotide molecules preferably encode NeuroD proteins that trigger differentiation in ectodermal cells, particularly neuroectodermal stem cells, and in more committed cells of that lineage, for example, epidermal precursor cells, pancreatic and gastrointestinal cells. Such neuroD expression products typically form heterodimeric bHLH protein complexes that bind in the 5'-regulatory regions of target genes and enhance or suppress transcription of the target gene.

In some instances, cancer cells may contain a nonfunctional NeuroD protein or may contain no NeuroD protein due to genetic mutation or somatic mutations such that these cells fail to differentiate. For cancers of this type, the cancer cells may be treated in a manner to cause the over-expression of wild-type NeuroD protein to force differentiation of the cancer cells.

Antisense neuroD nucleotide sequences may be used to block expression of mutant neuroD expression in neuronal precursor cells to generate and harvest neuronal stem cells. The use of antisense oligonucleotides and their applications have been reviewed in the literature (see, for example, Mol and Van der Krul, eds., *Antisense Nucleic Acids and Proteins Fundamentals and Applications*, New York, N.Y., 1992; which is incorporated by reference herein in its entirety). Suitable antisense oligonucleotides are at least 11 nucleotide in length and may include untranslated (upstream or intron) and associated coding sequences. As will be evident to one skilled in the art, the optimal length of an anti sense oligonucleotide depends on the strength of the interaction between the antisense oligonucleotide and the complementary mRNA, the temperature and ionic environment in which translation takes place, the base sequence of the antisense oligonucleotide, and the presence of secondary and tertiary structure in the mRNA and/or in the antisense oligonucleotide. Suitable target sequences for antisense oligonucleotides include intron-exon junctions (to prevent proper splicing), regions in which DNA/RNA hybrids will prevent transport of mRNA from the nucleus to the cytoplasm, initiation factor binding sites, ribosome binding sites, and sites that interfere with ribosome progression. A particularly preferred target region for antisense oligonucleotide is the 5' untranslated (promoter/enhancer) region of the gene of interest. Antisense oligonucleotides may be prepared by the insertion of a DNA molecule containing the target DNA sequence into a suitable expression vector such that the DNA molecule is inserted downstream of a promoter in a reverse orientation as compared to the gene itself. The expression vector may then be transduced, transformed or transfected into a suitable cell resulting in the expression of antisense oligonucleotides. Alternatively, antisense oligonucleotides may be synthesized using standard manual or automated synthesis techniques. Synthesized oligonucleotides may be introduced into suitable cells by a variety of means including electroporation, calcium phosphate precipitation, or microinjection. The selection of a suitable antisense oligonucleotide administration method will be evident to one skilled in the art. With respect to synthesized oligonucleotides, the stability of antisense oligonucleotide-mRNA hybrids may be increased by the addition of stabilizing agents to the oligonucleotide. Stabilizing agents include intercalating agents that are covalently attached to either or both ends of the oligonucleotide. Oligonucleotides may be made resistant to nucleases by, for example, modifications to the phosphodiester backbone by the introduction of phosphotriesters, phosphonates, phosphorothioates, phosphoroselenoates, phosphoramidates, or phosphorodithioates. Oligonucleotides may also be made nuclease resistant by synthesis of the oligonucleotides with alpha-anomers of the deoxyribonucleotides.

NeuroD binds to 5' regulatory regions of neurogenic genes that are involved in neuroectodermal differentiation, including development of neural and endocrine tissues. As described in more detail herein, murine neuroD has been detected in neuronal, pancreatic and gastrointestinal tissues in embryonic and adult mice suggesting that NeuroD functions in the transcription regulation in these tissues. NeuroD proteins alter the expression of subject genes by, for example, down-regulating or up-regulating transcription, or by inducing a change in transcription to an alternative open reading frame. The subject polynucleotide molecules find a variety of uses, e.g., in preparing oligonucleotide probes, expression vectors, and transformed host cells, as disclosed below in the following Examples.

DNA sequences recognized by NeuroD may be determined using a number of methods known in the literature including immunoprecipitation (Biedenkapp et al, *Nature* 335: 835–837, 1988; Kinzler and Vorgelstein, *Nuc. Acids Res.* 17: 3645–3653, 1989; and Sompayrac and Danna, *Proc. Natl. Acad Sci. USA* 87: 3274–3278, 1990; which are incorporated by reference herein), protein affinity columns (Oliphant et at., *Mol. Cell. Biol.* 9: 2944–2949, 1989; which is incorporated by reference herein), gel mobility shifts (Blackwell and Weintraub, *Science* 250: 1104–1110, 1990; which is incorporated by reference herein), and Southwestern blots (Keller and Maniatis, *Nuc. Acids Res.* 17:4675–4680, 1991; which is incorporated by reference herein).

One embodiment of the present invention involves the construction of interspecies hybrid NeuroD proteins and hybrid NeuroD proteins containing one or more domains from another NeuroD family member to facilitate structure-function analyses or to alter NeuroD activity by increasing or decreasing the transcriptional activation of neurogenic genes by NeuroD relative to the wild-type NeuroD(s). Hybrid proteins of the present invention may contain the replacement of one or more contiguous amino acids of the native NeuroD with the analogous amino acid(s) of NeuroD from another species or other protein of the NeuroD family. Such interspecies or interfamily hybrid proteins include hybrids having whole or partial domain replacements. Such hybrid proteins are obtained using recombinant DNA techniques. Briefly, DNA molecules encoding the hybrid NeuroD proteins of interest are prepared using generally available methods such as PCR mutagenesis, site-directed mutagenesis, and/or restriction digestion and ligation. The hybrid DNA is then inserted into expression vectors and introduced into suitable host cells. The biological activity may be assessed essentially as described in the assays set forth in more detail in the Examples that follow.

The invention also provides synthetic peptides, recombinantly derived peptides, fusion proteins, and the like that include a portion of NeuroD or the entire protein. The subject peptides have an amino acid sequence encoded by a nucleic acid which hybridizes under stringent conditions with an oligonucleotide of 15 or more contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:16. Representative amino acid sequences of the subject peptides are disclosed in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, and SEQ ID NO:17. The subject peptides find a variety of uses, including preparation of specific antibodies and preparation of antagonists of NeuroD activity.

As noted above, the invention provides antibodies that bind to NeuroD. The production of non-human antisera or monoclonal antibodies (e.g., murine, lagomorph, porcine, equine) is well known and may be accomplished by, for example, immunizing an animal with NeuroD protein or peptides. For the production of monoclonal antibodies, antibody producing cells are obtained from immunized animals, immortalized and screened, or screened first for the production of the antibody that binds to the NeuroD protein or peptides and then immortalized. It may be desirable to transfer the antigen binding regions (e.g., F(ab')2 or hypervariable regions) of non-human antibodies into the framework of a human antibody by recombinant DNA techniques to produce a substantially human molecule. Methods for producing such "humanized" molecules are generally well known and described in, for example, U.S. Pat. No. 4,816,397; which is incorporated by reference herein in its entirety. Alternatively, a human monoclonal antibody or portions thereof may be identified by first screening a human B-cell cDNA library for DNA molecules that encode antibodies that specifically bind to NeuroD, e.g., according to the method generally set forth by Huse et al. (*Science* 246: 1275–1281, 1989, which is incorporated by reference herein in its entirety). The DNA molecule may then be cloned and amplified to obtain sequences that encode the antibody (or binding domain) of the desired specificity.

The invention also provides methods for inducing the expression of genes associated with neuronal phenotype in a cell that does not normally express those genes. Examples of neuronal phenotypes that may be modulated by NeuroD expression include expression of neurotransmitters or neuromodulatory factors. Cells that can be used for the purpose of modulation of gene expression by NeuroD include cells of the neuroectodermal lineage, glial cells, neural crest cells, and epidermal epithelial basal stem cells, and all types of both mesodermal and endodermal lineage cells. NeuroD expression may also be used within methods that induce expression of genes associated with pancreatic and gastrointestinal phenotype. Examples of such gene expression include insulin expression, and gastrointestinal-specific enzyme expression.

As illustrated in Example 10, the expression of Xenopus NeuroD protein in stem cells causes redirection of epidermal cell differentiation and induces terminal differentiation into neurons, i.e., instead of epidermal cells. Epithelial basal stem cells (i.e., in skin and mucosal tissues) are one of the few continuously regenerating cell types in an adult mammal. Introduction of the subject nucleotide sequences into an epithelial basal stem cell may be accomplished in vitro or in vivo using a suitable gene therapy vector delivery system (e.g., a retroviral vector), a microinjection technique (see, for example, Tam, *Basic Life Sciences* 37: 187–194, 1986, which is incorporated by reference herein in its entirety), or a transfection method (e.g., naked or liposome encapsulated DNA or RNA; see, for example, *Trends in Genetics* 5: 138, 1989; Chen and Okayama, *Biotechniques* 6: 632–638, 1988; Mannino and Gould-Fogerite, *Biotechniques* 6: 682–690, 1988; Kojima et al., *Biochem. Biophys. Res. Comm.* 207: 8–12, 1995; which are incorporated by reference herein in their entirety). The introduction method may be chosen to achieve a transient expression of NeuroD in the host cell, or it may be preferable to achieve constitutive or regulated expression in a tissue specific manner.

Transformed host cells of the present invention find a variety of in vitro uses, for example: i) as convenient sources of neuronal and other growth factors, ii) in transient and continuous cultures for screening anti-cancer drugs capable of driving terminal differentiation in neural tumors, iii) as sources of recombinantly expressed NeuroD protein for use as an antigen in preparing monoclonal and polyclonal antibodies useful in diagnostic assays, and iv) in transient and continuous cultures for screening for compounds capable of increasing or decreasing the activity of NeuroD.

Transformed host cells of the present invention also find a variety of in vivo uses, for example, for transplantation at sites of traumatic neural injury where motor or sensory neural activity has been lost. Representative patient populations that may benefit from transplantation include: patients with hearing or vision loss due to optical or auditory nerve damage, patients with peripheral nerve damage and loss of motor or sensory neural activity, and patients with brain or spinal cord damage from traumatic injury. For example, donor cells from a patient such as epithelial basal stem cells are cultured in vitro and then transformed or transduced with a neuroD nucleotide sequence. The transformed cells are then returned to the patient by microinjection at the site of neural dysfunction. In addition, transformed host cells of the present invention may be useful for transplantation into patients with diabetes. For example, donor cells from a patient such as fibroblasts, pancreatic islet cells, or other pancreatic cells are harvested and transformed or transfected with a neuroD nucleotide sequence. The genetically engineered cells are then returned to the patient. In another embodiment, such engineered host cells may find use in the treatment of malabsorption syndromes.

Representative uses of the nucleotide sequences of the invention include the following:

1. Construction of cDNA and oligonucleotide probes useful in Northern, Southern, and dot-blot assays for iden-

tifying and quantifying the level of expression of neuroD in a cell. High level expression of neuroD in neuroendocrine tumors and in rapidly proliferating regions of embryonic neural development (see below) indicates that measuring the level of neuroD expression may provide prognostic markers for assessing the growth rate and invasiveness of a neural tumor. In addition, considering the important role of NeuroD in embryonic development it is thought highly likely that birth defects and abortions may result from expression of an abnormal NeuroD protein. In this case, NeuroD may prove highly useful in prenatal screening of mothers and/or for in utero testing of fetuses.

2. Construction of recombinant cell lines, ova, and transgenic embryos and animals including dominant-negative and "knock-out" recombinant cell lines in which the transcription regulatory activity of NeuroD protein is down-regulated or eliminated. Such cells may contain altered neuroD coding sequences that result in the expression of a NeuroD protein that is not capable of enhancing, suppressing or activating transcription of the target gene. The subject cell lines and animals find uses in screening for candidate therapeutic agents capable of either substituting for a function performed by NeuroD or correcting the cellular defect caused by a defective NeuroD. Considering the important regulatory role of NeuroD in embryonic development, birth defects may occur from expression of mutant NeuroD proteins, and these defects may be correctable in utero or in early post-natal life through the use of compounds identified in screening assays using NeuroD. In addition, neuroD polynucleotide molecules may be joined to reporter genes, such as β-galactosidase or luciferase, and inserted into the genome of a suitable embryonic host cell such as a mouse embryonic stem cell by, for example, homologous recombination (for review, see Capecchi, *Trends in Genetics* 5: 70–76, 1989; which is incorporated by reference). Cells expressing NeuroD may then be obtained by subjecting the differentiating embryonic cells to cell sorting, leading to the purification of a population of neuroblasts. Neuroblasts may be useful for studying neuroblast sensitivity to growth factors or chemotherapeutic agents. The neuroblasts may also be used as a source from which to purify specific protein products or gene transcripts. These products may be used for the isolation of growth factors, or for the identification of cell surface markers that can be used to purify stem cell population from a donor for transplantation.

As illustrated in Example 14, "knock-out" mice were generated by replacing the murine neuroD coding region with the β-galactosidase reporter gene and the neomycin resistance gene to assess the consequences of eliminating the murine NeuroD protein and to examine the tissue distribution of NeuroD in fetal and postnatal mice. Mice that were homozygous for the mutation (lacking NeuroD) had diabetes, as demonstrated by high blood glucose levels, and died by day four. Homozygous mutants had blood glucose levels between 2 and 3 times the blood glucose level of wild-type mice. Heterozygous mutants exhibited similar blood glucose levels as wild-type mice. Examination of stained tissue from fetal and postnatal mice heterozygous for the mutation confirmed the NeuroD expression pattern in neuronal cells demonstrated by in situ hybridization (Example 4) and also demonstrated neuroD expression in the pancreas and gastrointestinal tract.

"Knock-out" mice may be useful as a model system for diabetes. Such mice may be used to study methods to rescue homozygous mutants and as hosts to test transplant tissue for treating diabetes.

3. Construction of gene transfer vectors (e.g., retroviral vectors, and the like) wherein neuroD is inserted into the coding region of the vector under the control of a promoter. neuroD gene therapy may be used to correct traumatic neural injury that has resulted in loss of motor or sensory neural function, and also for the treatment of diabetes. For these therapies, gene transfer vectors may either be injected directly at the site of the traumatic injury, or the vectors may be used to construct transformed host cells that are then injected at the site of the traumatic injury. The results disclosed in Example 10 indicate that introduction of neuroD induces a non-neuronal cell to become a neuron. This discovery raises for the first time the possibility of using transplantation and/or gene therapy to repair neural defects resulting from traumatic injury. In addition, the discovery of neuroD provides the possibility of providing specific gene therapy for the treatment of certain neurological disorders such as Alzheimer's disease, Huntington's disease, and Parkinson's disease, in which a population of neurons have been damaged. Two basic methods of neuroD utilization are envisioned in this regard. In one method, neuroD is expressed in existing populations of neurons to modulate aspects of their neuronal phenotype (e.g., neurotransmitter expression or synapse targeting) to make the neurons express a factor or phenotype to overcome the deficiency that contributes to the disease. In this method, recombinant neuroD sequences are introduced into existing neurons or endogenous neuroD expression is induced. In another method, neuroD is expressed in non-neuronal cells (e.g., glial cells in the brain or another non-neuronal cell type such as basal epithelial cells) to induce expression of genes that confer a complete or partial neuronal phenotype that ameliorates aspects of the disease. As an example, Parkinson's disease is caused, at least in part, by the death of neurons that supply the neurotransmitter dopamine to the basal ganglia. Increasing the levels of neurotransmitter ameliorates the symptoms of Parkinson's disease. Expression of neuroD in basal ganglia neurons or glial cells may induce aspects of a neuronal phenotype such that the neurotransmitter dopamine is produced directly in these cells. It may also be possible to express neuroD in donor cells for transplantation into the affected region, either as syngeneic or allogeneic transplantations. Within yet another embodiment, neuroD is expressed in non-pancreatic cells to induce expression of genes that confer a complete or partial pancreatic phenotype that ameliorates aspects of diabetes. Within yet another embodiment, neuroD is expressed in pancreatic islet cells to induce expression of genes that induce the expression of insulin.

4. Preparation of transplantable recombinant neuronal precursor cell populations from embryonic ectodermal cells, non-neural basal stem cells, and the like. Establishing cultures of non-malignant neuronal cells for use in therapeutic screening assays has proven to be a difficult task. The isolated polynucleotide molecules encoding NeuroD of the present invention permit the establishment of primary (or continuous) cultures of proliferating embryonic neuronal stem cells under conditions mimicking those that are active in development and cancer. The resultant cell lines find uses: i) as sources of novel neural growth factors, ii) in screening assays for anti-cancer compounds, and iii) in assays for identifying novel neuronal growth factors. High level expression of neuroD in the embryonic optic rectum (see below) indicates that NeuroD protein may regulate expression of factors trophic for growing retinal cells. Such cells may be useful sources of growth factors, and may be useful in screening assays for candidate therapeutic compounds.

The cell lines and transcription regulatory factors disclosed herein offer the unique advantage that since they are

active very early in embryonic differentiation they represent potential switches, e.g., ON→OFF or OFF→ON, controlling subsequent cell fate. If the switch can be shown to be reversible (i.e., ON↔OFF), the NeuroD transcription regulatory factor and neuroD nucleic acids disclosed herein provide exciting opportunities for restoring lost neural and/or endocrine functions in a subject.

The following examples are offered by way of illustration and not by way of limitation.

EXAMPLE 1

Construction of the embryonic stem cell "179" cDNA libray

A continuous murine embryonic stem cell line (i.e., the ES cell line) having mutant E2A (the putative binding partner of myoD) was used as a cell source to develop a panel of embryonic stem cell tumors. Recombinant ES stem cells were constructed (i.e., using homologous recombination) wherein both alleles of the putative myoD binding partner E2A were replaced with drug-selectable marker genes. ES cells do not make functional E12 or E47 proteins, both of which are E2A gene products. ES cells form subcutaneous tumors in congenic mice (i.e., 129J) that appear to contain representatives of many different embryonal cell types as judged histologically and through the use of RT-PCR gene expression assays. Individual embryonic stem cell tumors were induced in male 129J strain mice by subcutaneous injection of 1×10^7 cells/site. Three weeks later each tumor was harvested and used to prepare an individual sample of RNAs. Following random priming and second strand synthesis the ds-cDNAs were selected based on their size on 0.7% agarose gels and those cDNAs in the range of 400–800 bp were ligated to either Bam HI or Bgl H linkers. (Linkers were used to minimize the possibility that an internal Bam HI site in a cDNA might inadvertently be cut during cloning, leading to an abnormally sized or out-of-frame expression product.) The resultant individual stem cell tumor DNAs were individually ligated into the Bam HI cloning site in the "fl-VP16" 2μ yeast expression vector. This expression vector, fl-VP16, contains the VP16 activation domain of Herpes simplex virus (HSV) located between Hind III (HIII) and Eco RI (RI) sites and under the control of the *Saccharomyces cereviseae* alcohol dehydrogenase promoter; with LEU2 and Ampicillin-resistance selectable markers. Insertion of a DNA molecule of interest into the Hind HI site of the fl-VP 16 vector (i.e., 5' to the VP16 nucleotide sequence), or into a Bam HI site (i.e., 3' to the VP16 sequence but 5' to the Eco RI site), results in expression of a VP16 fusion protein having the protein of interest joined in-frame with VP 16. The resultant cDNA library was termed the "179-library".

EXAMPLE 2

Identification and cDNA cloning of neuroD

A two-hybrid yeast screening assay was used essentially as described by Fields and Song (Nature 340: 245, 1989) and modified as described herein was used to screen the 179-library described in Example 1. Yeast two-hybrid screens are reviewed as disclosed in Fields and Sternglanz (*Trends in Genetics* 10: 286–292, 1994). The library was screened for cDNAs that interacted with LexA-Da, a fusion protein between the Drosophila Da (Daughterless) bHLH domain and the prokaryotic LexA-DNA binding domain. Multimerized LexA binding sites were cloned upstream of two reporter genes, the HIS3 gene and the β-galactosidase gene. The *S. cereviseae* strain L40 containing a plasmid encoding the LexA-Da fusion protein was transformed with CsCl gradient-purified fl-VP16-179-cDNA library. Transformants were maintained on medium selecting both plasmids (the LexA-Da plasmid and the cDNA library plasmid) for 16 hours before being subjected to histidine selection on plates lacking histidine, leucine, tryptophan, uracil, and lysine. Clones that were HIS^+ were subsequently assayed for the expression of LacZ. To eliminate possible non-specific cloning artifacts, plasmids from $HIS^+/LacZ^+$ were isolated and transformed into *S. cereviseae strain L40* containing a plasmid encoding a LexA-Lamin fusion. Clones that scored positive in the interaction with lamin were discarded. Approximately 400 cDNA clones, which represented 60 different transcripts, were identified as positive in these assays. Twenty-five percent of the original clones were subsequently shown to be known bHLH genes on the basis of their reactivity with specific cDNA probes. One cDNA clone encoding a VP16-fusion protein that interacted with Da but not lamin was identified as unique by sequence analysis. This clone, initially termed tango, is now referred to as neuroD.

The unique cDNA identified above, VP 16-neuroD, contained an approximately 450 bp insert that spanned the bHLH region. Sequence analysis showed that the clone contained an insert encoding a complete bHLH amino acid sequence motif that was unique and previously unreported. Further analysis suggested that while the cDNA contained conserved residues common to all members of the bHLH protein family, several residues were unique and made it distinct from previously identified bHLH proteins. The neuroD cDNA insert was subcloned as a Bam HI-Not I insert into Bam HI-Not I linearized pBluescript SK^+. The resulting plasmid was designated pSK+1-83.

The neuroD insert contained in the VP16-neuroD plasmid was used to re-probe a mouse cDNA library prepared from mouse embryos at developmental stage e10.5. Candidate clones were isolated and sequenced essentially as described above. Several clones were isolated. One clone, designated pKS^+m7a RX, was deposited at the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Md. 20852 USA, on May 6, 1994, under accession number 75768. Plasmid pKS m7a RX contains 1646 bp of murine neuroD cDNA as an EcoRI-XhoI insert. The amino acid sequence encoded by the insert begins at amino acid residue +73 and extends to the carboxy-terminus of the NeuroD protein. The plasmid contains about 855 bp of NeuroD coding sequence. (encoding amino acids 73–536).

None of the mouse cDNAs contained the complete 5' coding sequence. To obtain the 5' neuroD coding sequence, a mouse strain 129/Sv genomic DNA library was screened with the VP16-neuroD plasmid insert (450 bp). Genomic clones were isolated and sequenced and the sequences were aligned with the cDNA sequences. Alignment of the sequence and comparison of the genomic 5' coding sequences with the Xenopus neuroD clone (Example 8) confirmed the 5' neuroD coding sequence. The complete neuroD coding sequence and deduced amino acid sequence are shown in SEQ ID NOS: 1 and 2.

EXAMPLE 3

NeuroD/neuroD

bHLH proteins share common structural similarities that include a basic region that binds DNA and an HLH region

involved in protein-protein interactions required for the formation of homodimers and heterodimeric complexes. A comparison of the amino add sequence of the basic region of murine NeuroD (amino acids 102 to 113 of SEQ ID NO:2) with basic regions of other bHLH proteins revealed that murine NeuroD contained all of the conserved residues characteristic among this family of proteins. However, in addition, NeuroD contained several unique residues. These unique amino acid residues were not found in any other known HLH making NeuroD a distinctive new member of the bHLH family. The NARERNR basic region motif in NeuroD (amino adds 107–113 of SEQ ID NO:2) is also found in the Drosophila AS-C protein, a protein thought to be involved in neurogenesis. Similar, but not identical, NARERRR and NERERNR motifs (SEQ ID NOS:5 and 6, respectively) have been found in the Drosophila Atonal and MASH (mammalian achaete-scute homolog) proteins, respectively, which are also thought to be involved in neurogenesis. The NARER motif (SEQ ID NO:7) of neuroD is shared by other bHLH proteins, and the Drosophila Daughterless (Da) and Mammalian E proteins. The basic region of bHLH proteins is important for DNA binding site recognition, and there is homology between NeuroD and other neuro-proteins in this functional region. Within the important dimer-determining HLH region of NeuroD, a low level of homology was recorded with mouse twist protein (i.e., 51% homology) and with MASH (i.e., 46% homology). NeuroD contains several regions of unique peptide sequence within the bHLH domain including the junction sequence (MHG).

EXAMPLE 4

NeuroD is expressed in differentiating neurons during embryonic development

neuroD expression was analyzed during embryonic development of mouse embryos using in situ hybridization with an antisense neuroD single-stranded riboprobe labeled with digoxigenin (Boehringer Mannheim). Briefly, a riboprobe was prepared from plasmid pSK+1-83 using T7 polymerase and digoxigenin-11-UTP for labeling. The hybridized probe was detected using anti-digoxigenin antibody conjugated with alkaline phosphatase. Color development was carried out according to the manufacturer's instruction. Stages of development are commonly expressed as days following copulation and where formation of the vaginal plug is e0.5. The results recorded in the in situ hybridization studies were as follows:

In the e9.5 mouse embryo, neuroD expression was observed in the developing trigerminal ganglia.

In the e10.5 mouse embryo, a distinctive pattern of neuroD expression was observed in all the cranial ganglia (i.e., V–XI) and in dorsal root ganglia (DRG) in the trunk region of the embryo. At this time neuroD expression was also observed in the central nervous system in post-mitotic cells in the brain and spinal cord that were undergoing neuronal differentiation. In the spinal cord, the ventral portion of the cord from which the motor neurons arise and differentiate was observed to express neuroD at high levels; and expression in the posterior-ventral spinal cord was higher when compared to more mature anterior-ventral spinal cord.

In the e11.5 mouse embryo, the ganglionic expression pattern of neuroD observed in e10.5 persisted. Expression in the spinal cord was increased over the level of expression observed in e10.5 embryos, which is consistent with the presence of more differentiating neurons at this stage. At this stage neuroD expression is also observed in other sensory organs in which neuronal differentiation occurs, for example, in the nasal epithelium, otic vesicle, and retina of the eye. In both of these organs neuroD expression was observed in the region containing differentiating neurons.

In the e14.5 mouse embryo, expression of neuroD was observed in cranial ganglia and DRG, but expression of neuroD persisted in the neuronal regions of developing sensory organs and the central nervous system (CNS). Thus, neuroD expression was observed to be transient during neuronal development.

In summary, expression of neuroD in the neurula stage of the embryo (e10), in the neurogenic derivatives of neural crest cells, the cranial and dorsal root ganglia, and post mitotic cells in the CNS suggests an important possible link between expression and generation of sensory and motor nerves. Expression occurring later in embryonic development in differentiating neurons in the CNS and in sensory organs (i.e., nasal epithelium and retina) also supports a role in development of the CNS and sensory nervous tissue. Since neuroD expression is transient, the results suggest that neuroD expression is operative as a switch controlling formation of sensory nervous tissue. It is noteworthy that in these studies neuroD expression was not observed in embryonic sympathetic and enteric ganglia (also derived from migrating neural crest cells). Overall, the results indicate that neuroD plays an important role in neuronal differentiation.

EXAMPLE 5

NeuroD is expressed in neural and brain tumor cells: murine probes identify human neuroD

Given the expression pattern in mouse embryo (Example 4), Northern blots of tumor cell line mRNAs were examined using murine neuroD cDNA (Example 2) as a molecular probe. As a first step, cell lines that have the potential for developing into neurons were screened. The D283 human medullablastoma cell line, which expressed many neuronal markers, expressed high levels of neuroD by Northern blot analysis. neuroD was also transcribed at various levels by different human neuroblastoma cell lines and in certain rhabdomyosarcoma lines that are capable of converting to neurons. Murine PC12 pheochromacytoma cells and P19 embryocarcinoma cells differentiate into neurons in tissue culture in the presence of appropriate inducers, i.e., nerve growth factor and retinoic acid, respectively. When induced, murine P19 but not PC12 cells expressed neuroD transcripts. However, non-induced murine PC12 cells, P19 cells, and control 3T3 fibroblasts did not produce detectable levels of neuroD transcripts. Thus, PC12 and P19 cells represent cell types that are potentially useful in screening assays for identifying inducers of neuroD expression that may stimulate nerve regeneration and differentiation of neural tumor cells.

EXAMPLE 6

Recombinant cells expressing NeuroD

Recombinant murine 3T3 fibroblast cells expressing either a myc-tagged murine NeuroD protein or myc-tagged Xenopus NeuroD protein were made. The recombinant cells were used as a test system for identifying antibody to NeuroD described below.

Xenopus NeuroD protein was tagged with the antigenic marker Myc to allow the determination of the specificity of

anti-NeuroD antibodies to be determined. Plasmid CS2+MT was used to produce the Myc fusion protein. The CS2+MT vector (Turner and Weintraub, ibid.) contains the simian cytomegalovirus IE94 enhance/promoter (and an SP6 promoter in the 5' untranslated region of the IE94-driven transcript to allow in vitro RNA synthesis) operatively linked to a DNA sequence encoding six copies of the Myc epitope tag (Roth et al, *J. Cell Biol.* 115: 587–596, 1991; which is incorporated herein in its entirety), a polylinker for insertion of coding sequences, and an SV40 late polyadenylation site. CS2-MT was digested with Xho I to linearize the plasmid at the polylinker site downstream of the DNA sequence encoding the myc tag. The linearized plasmid was blunt-ended using Klenow and dNTPs. A full length Xenopus cDNA clone was digested with Xho I and Eae I and blunt-ended using Klenow and dNTPs, and the 1.245 kb fragment of the Xenopus neuroD cDNA was isolated. The neuroD fragment and the linearized vector were ligated to form plasmid CS2+MT x1-83.

CS2+MT was digested with Eco RI to linearize the plasmid at the polylinker site downstream of the DNA sequence encoding the myc tag. The linearized plasmid was blunt-ended using Klenow and dNTPs and digested with Xho I to obtain a linearized plasmid having an Xho I adhesive end and a blunt end. Plasmid pKS+m7a containing a partial murine neuroD cDNA was digested with Xho I, and the NeuroD containing fragment was blunt-ended and digested with Xba I to obtain the approximately 1.6 kb fragment of the murine neuroD cDNA. The neuroD fragment and the linearized vector were ligated to form plasmid CS2+MT M1–83(m7a).

Plasmids CS2+MT x1-83 and CS2+MT M1-83(m7a) were each transformed into murine 3T3 fibroblast cells and used as a test system for identifying antibody against NeuroD (Example 7).

EXAMPLE 7

Antibodies to NeuroD

A recombinant fusion protein of maltose binding protein (MBP) and amine acid residues 70–355 of murine NeuroD was used as an antigen to evoke antibodies in rabbits. Specificity of the resultant antisera was confirmed by immunostaining of the recombinant 3T3 cells described above. Double-immunostaining of the recombinant cells was observed with monoclonal antibodies to Myc (i.e., the control antigenic tag on the transfected DNA) and with rabbit anti-murine NeuroD in combination with anti-rabbit IgG. The specificity of the resultant anti-murine NeuroD sera was investigated further by preparing mouse 3T3 fibroblasts cells transfected with different portions of NeuroD DNA. Specificity seemed to map to the glutamic acid-rich domain (i.e., amine acids 66–73 of SEQ ID NO:2). The anti-murine antisera did not react with cells transfected with the myc-tagged Xenopus neuroD. In a similar manner, Xenopus NeuroD was used to generate rabbit anti-NeuroD antisera. The antisera was Xenopus-specific and did not cross react with cells transfected with myc-tagged murine neuroD.

EXAMPLE 8

NeuroD is a highly evolutionarily conserved protein: sequence of Xenopus NeuroD

Approximately one million clones from a stage 17 Xenopus head library made by Kintner and Melton (Development 99: 311, 1987) were screened with the mouse cDNA insert as a probe at low stringency. The hybridization was performed with 50% formamide/4× SSC at 33° C. and washed with 2× SSC/0.1% SDS at 40° C.

Positive clones were identified and sequenced. Analysis of the Xenopus neuroD cDNA sequence (SEQ ID NO:3) revealed that NeuroD is a highly conserved protein between frog and mouse. The deduced amino acid sequences of frog and mouse (SEQ ID NOS:2 and 4) show 96% identity in the bHLH domain (50 of 52 amino acids are identical) and 80% identity in the region that is carboxy-terminal to the bHLH domain (159 of 198 amino acids are identical). The domain structures of murine and Xenopus NeuroD are highly homologous with an "acidic" N-terminal domain (i.e., glutamic or aspartic acid rich); a basic region; helix 1, loop, helix 2; and a proline rich C-terminal region. Although the amino terminal regions of murine and Xenopus NeuroD differ in amino acid sequence, both retain a glutamic or aspartic acid rich "acidic domain" (amino acids 102 to 113 of SEQ ID NO:2 and amino acids 56 to 79 of SEQ ID NO:4). It is highly likely that the acidic domain constitutes an "activation" domain for the NeuroD protein, in a manner analogous to the activation mechanisms currently understood for other known transcription regulatory factors.

EXAMPLE 9

Neuronal expression of Xenopus neuroD

The expression pattern of neuroD in whole mount Xenopus embryos was determined using in situ hybridization with a single stranded digoxigenin-labeled Xenopus neuroD antisense cDNA riboprobe. Embryos were examined at several different stages.

Consistent with the mouse expression pattern, by late stage, all cranial ganglia showed very strong staining patterns. In Xenopus, as in other vertebrate organisms, neural crest cells give rise to skeletal components of the head, all ganglia of the peripheral nervous system, and pigment cells. Among these derivatives, the cranial sensory ganglia, which are of mixed crest and placode origin, represent the only group of cells that express neuroD. High levels of neuroD expression in the eye were also observed, correlating with active neuronal differentiation in the retina at this stage. Expression is observed in the developing olfactory placodes and otic vesicles, as was seen in mice. The pineal gland also expressed neuroD. All of this expression in transient, suggesting that neuroD functions during the differentiation process but is not required for maintenance of these differentiated cell types.

As early as stage 14 (i.e., the mid-neurula stage) neuroD expression was observed in the cranial neural crest region where trigerminal ganglia differentiate. Primary mechanosensory neurons in the spinal cord, also referred to as Rohon-Beard cells and primary motor neurons, showed neuroD expression at this stage.

By stage 24, all of the developing cranial ganglia, trigerminal, facio-acoustic, glosso-pharyngeal, and vagal nervous tissues showed a high level of neuroD expression. High levels of expression of neuroD was also observed in the eye at this stage. (Note that in Xenopus neuronal differentiation in the retina occurs at a much earlier stage than in mice, and neuroD expression was correspondingly earlier and stronger in this animal model.)

In summary, in Xenopus as in mouse, neuroD expression was correlated with sites of neuronal differentiation. The remarkable evolutionary conservation of the pattern of neu-

roD expression in differentiating neurons supports the notion that NeuroD has been evolutionarily conserved both structurally and functionally in these distant classes, which underscores the critical role performed by this protein in embryonic development.

EXAMPLE 10

Ectopic expression of neuroD converts non-neuronal cells into neurons

To further analyze the biological functions of NeuroD, a gain-of-function assay was conducted. In this assay, RNA was microinjected into one of the two cells in a 2-cell stage Xenopus embryo, and the effects on later development of neuronal phenotype was evaluated. For these experiments myc-tagged neuroD transcripts were synthesized in vitro using SP6 RNA polymerase. The myc tagged-neuroD transcripts were microinjected into one of the two cells in a Xenopus 2-cell embryo, and the other cell of the embryo served as an internal control. Antibodies to Xenopus N-CAM, a neural adhesion molecule, anti-Myc (to detect the exogenous protein), and immunostaining techniques were used to evaluate phenotypic expression of the neuronal marker (and control) gene during the subsequent developmental stages of the microinjected embryos. Remarkably, an evaluation of over 130 embryos that were injected with neuroD RNA showed a striking increase in ectopic expression of N-CAM on the microinjected side of the embryo (i.e., Myc^+), as judged by increased immunostaining. The increased staining was observed in the region from which neural crest cells normally migrate. It is considered likely that ectopic expression (or over-expression) of neuroD caused neural crest stem cells to follow a neurogenic cell fate. Outside the neural tube, the ectopic immunostaining was observed in the facio-cranial region and epidermal layer, and in some cases the stained cells were in the ventral region of the embryo far from the neural tube. The immunostained cells not only expressed N-CAM ectopically, but displayed a morphological phenotype of neuronal cells. At high magnification, the N-CAM expressing cells exhibited typical neuronal processes reminiscent of axonal processes.

To confirm that the ectopic N-CAM expression resulted from a direct effect on the presumptive epidermal cells and not from aberrant neural cell migration into the lateral and ventral epidermis, neuroD RNA was injected into the top tier of 32-cell stage embryos, in order to target the injection into cells destined to become epidermis. N-CAM gaining was observed in the lateral and ventral epidermis without any noticeable effect on the endogenous nervous system, indicating that the staining of N-CAM in the epidermis represents the conversion of epidermal cell fate into neuronal cell fate.

Ectopic generation of neurons by neuroD was confirmed with other neural specific markers, such as neural-specific class II β-tubulin (Richter et at., *Proc. Natl. Acad. Sci. USA* 85: 8066, 1988), acetylated alpha-tubulin (Piperno and Fuller, *J. Cell. Biol.* 101: 2085, 1985), tanabin (Hemmati-Brinvanlou et at., *Neuron* 9: 417, 1992), neurofilament (NF)-M (Szaro et at., *J. Comp. Neurol.* 273: 344, 1988), and Xen-1,2 (Ruiz i Altaba, *Development* 115: 67, 1992). The embryos were subjected to immunochemistry as described by Turner and Weintraub (*Genes Dev.* 8:1434, 1994, which is incorporated by reference herein) using primary antibodies detected with alkaline phosphatase-conjugated goat anti-mouse or anti-rabbit antibodies diluted to 1:2000 (Boehringer-Mannheim). Anti-acetylated alpha-tubulin was diluted 1:2000. Anti-Xen-1 was diluted 1:1. Anti-NF-M was diluted 1:2000. Embryos stained for NF-M were fixed in Dent's fixative (20% dimethylsulfoxide/80% methanol) and cleared in 2:1 benzyl benzoate/benzyl alcohol as described by Dent et at. (*Development* 105: 61, 1989, which is incorporated by reference herein). In situ hybridization of embryos was carried out essentially as described by Harland (in *Methods in Cell Biology*, B. K. Kay, H. J. Pend, Eds, Academic Press, New York, N.Y., Vol 36, pp. 675–685, 1991, which is incorporated by reference herein) as modified by Turner and Weintraub (ibid.). In situ hybridization with β-tubulin without RNase treatment can also detect tubulin expression in the ciliated epidermal cells. All of these markers displayed ectopic sting on the neuroD RNA injected side. Injection of neuroD mRNA into vegetal cells led to no ectopic expression of neural markers except in one embryo that showed internal N-CAM staining in the trunk region, suggesting the absence of cofactors or the presence of inhibitors in vegetal cells. However, the one embryo that showed ectopic neurons in the internal organ tissue suggests that it may be possible to convert non-ectodermal lineage cells into neurons under certain conditions.

The embryos were also stained with markers that detect Rohon-Beard cells (cells in which neuroD is normally expressed). Immunostaining using the method described above for Rohon-Beard cell-specific markers such as HNK-1 (Nordlander, *Dev. Brain Res.* 50: 147, 1989, which is incorporated by reference herein) at a dilution of 1:1, Islet-1 (Ericson et al., *Science* 256: 1555, 1992 and Korzh et al., *Development* 118: 417, 1993) at a dilution of 1:500, and in situ hybridization as described above with shaker-1 (Ribera et al., *J. Neurosci.* 13: 4988, 1993) showed more cells staining on the injected side of the embryos.

The combined results support the notion that ectopic expression of NeuroD induced differentiation of neuronal cells from cells that, without neuroD microinjection, would have given rise to non-neuronal cells. In summary, these experiments support the notion that ectopic neuroD expression can be used to convert a non-neuronal cell (i.e., uncommitted neural crest cells and epidermal epithelial basal stem cells) into a neuron. These findings offer for the first time the potential for gene therapy to induce neuron formation in injured neural tissues.

Interesting morphological abnormalities were observed in the microinjected embryos. In many cases the eye on the microinjected side of the embryo failed to develop. In other embryos, the spinal cord on the microinjected side of the embryo failed to develop properly, and the tissues were strongly immunopositive when stained with anti-N-CAM. In addition, at the mid-neurula stage many microinjected embryos exhibited an increase in cell mass in the cranial region of the embryo from which (in a normal embryo) the neural crest cells and their derivatives (i.e., cranial ganglionic cells) would migrate. The observed cranial bulge exhibited strong immunostaining with antibodies specific for N-CAM. These results were interpreted to mean that morphological changes in the eye, neural crest, and spinal cord resulted from premature neural differentiation which altered the migration of neural and neural crest precursor cells.

NeuroD-injected embryos were also assayed for alteration in the expression of Xtwist, the Xenopus homolog of Drosophila twist, to determine whether neuroD converted non-neuronal components of neural crest cells into the neural lineage. In wild-type embryos, Xtwist is strongly expressed in the non-neuronal population cephalic neural crest cells that give rise to the connective tissue and skeleton of the head. neuroD-injected embryos were completely missing Xtwist expression in the migrating cranial neural crest cells

on the injected side. The failure to generate sufficient cranial mesenchymal neural crest precursors in neuroD-injected embryos was also observed morphologically, since many of the injected embryos exhibited poor branchial arch development in the head. Furthermore, the increased mass of cells in the cephalic region stained very strongly for N-CAM, β-tubulin, and Xen-1, indicating that these cells were neural in character.

The converse experiment in which frog embryos were injected with Xtwist mRNA showed that ectopic expression of Xtwist significantly decreased neuroD expression on the injected side. Thus, two members of the bHLH family, neuroD and Xtwist, may compete for defining the identity of different cell types derived from the neural crest. In the neuroD-injected embryos, exogenous neuroD may induce premigratory neural crest to differentiate into neurons in situ, and consequently they fail to migrate to their normal positions.

The effect of introduction of exogenous neuroD on the fate of cells that normally express neuroD, such as cranial ganglia, eye, otic vesicle, olfactory organs, and primary neurons, and on other CNS cells that normally do not express neuroD, was determined by staining for differentiation markers. When the cranial region of the embryo is severely affected by ectopic neuroD, the injected side of the embryos displayed either small or no eyes in addition to poorly organized brains, otic vesicles, and olfactory organs. Moreover, as the embryos grew, the spinal cord showed retarded growth, remaining thinner and shorter on the neuroD-injected side.

N-CAM staining in the normal embryo at early stages was not uniform throughout the entire neural plate, but rather was more prominent in the medial region of the neural plate. Injected embryos analyzed for N-CAM expression show that the neural plate on the injected side of the early stage embryos was stained more intensely and more laterally. The increase in N-CAM staining was not associated with any lateral expansion of the neural plate as assayed by visual inspection and staining with the epidermal marker EpA. This was in contrast to what has been observed with XASH-3 injection that causes neural plate expansion. These observations suggest that the first effects of neuroD are to cause neuronal precursors in the neural plate to differentiate prematurely.

To determine whether neuroD caused neuronal precursors to differentiate prematurely, injected embryos were stained using two neuronal markers that are expressed in differentiated neurons, neural specific β-tubulin and tanabin. In situ hybridization for β-tubulin and tanabin was carried out as described above. Over-expression of neuroD dramatically increased the β-tubulin signals in the region of the neural plate containing both motor neurons and Rohon-Beard cells at stage 14. The earliest ectopic β-tubulin positive cells on the injected side were observed at the end of gastrulation when the control side did not yet show any β-tubulin positive cells. Tanabin was also expressed in more cells in the spinal cord in the neuroD injected side of the embryos at stage 14. These results suggest that neuroD can cause premature differentiation of the neural precursors into differentiated neurons. This is a powerful indication that, when ectopically expressed or over-expressed, NeuroD can differentiate mitotic cells into non-dividing mature neurons.

EXAMPLE 11

Human genomic clones of neuroD, neuroD2 and neuroD3

Genomic clones encoding human NeuroD were obtained by probing a human fibroblast genomic library with the mouse neuroD cDNA. Host *E coli* strain LE392 (New England Biolabs) were grown in LB+10 mM $MgSO_4$ 0.2% maltose overnight at 37° C. The cells were harvested and resuspended in 10 mM $MgSO_4$ to a final OD600 of 2. The resuspended cells were used as hosts for phage infection. The optimal volume of phage stock for use in this screening was determined by using serial dilutions of the phage stock of a human fibroblast genomic library in lambda FIX II (Stratagene) to infect LE392 cells (New England Biolabs). To obtain approximately 50,000 plaques per plate, a 2.5 μl aliquot of the phage stock was used to infect 600 μl of the resuspended LE392 cells. The cells were incubated with the phage for 15 minutes at 37° C, after which the cells were mixed with 6.5 ml of top agar warmed to 50° C. The top agar was plated on solid LB, and incubated overnight at 37° C. A total of 22 15-cm plates were prepared in this manner.

Duplicate plaque lifts were prepared. A first set of Hybond membranes (Amersham) were placed onto the plates and allowed to sit for 2 minutes. The initial membranes were removed and the duplicate membranes were laid on the plates for 4 minutes. The membranes were allowed to air dry; then the phage were denatured in 0.5M NaOH, 1.5M NaCl for 7 minutes. The membranes were neutralized with two washes in neutralization buffer (1.5M NaCl, 0.5M Tris, pH 7.2). After neutralization, the membranes were crosslinked by exposure to UV. A 1 kb Eco RI-Hind III fragment containing murine neuroD coding sequences was random primed using the Random Priming Kit (Boehringer Mannheim) according to the manufacturer's instructions. Membranes were prepared for hybridization by placing six membranes in 10 ml of FBI hybridization buffer [100 g polyethylene glycol 800, 350 ml 20% SDS, 75 ml 20× SSPE; add water to a final volume of one liter] and incubating the membranes at 65° C. for 10 minutes. After 10 minutes, denatured salmon sperm DNA was added to a final concentration of 10 μg/ml and denatured probe was added to a final concentration of $0.25–0.5\times10^7$ cpm/ml. The membranes were hybridized at 65° C. for a period of 8 hours to overnight. After incubation, the SDS for 30 minutes at 50° C. The first wash was followed by a final wash in 0.1× SSC, 0.1% SDS for 30 minutes at 55° C. Autoradiographs of the membranes were prepared. The first screen identified 55 putative positive plaques. Thirty-one of the plaques were subjected to a secondary screen using the method essentially set forth above. Ten positive clones were identified and subjected to a tertiary screen as described above. Eight positive clones were identified after the tertiary screen. Of these eight clones, three (14B1, 9F1 and 20A1) were chosen for further analysis. Clones 14B1 and 20A1 were deposited at the American Type Culture Collection, 1: 2301 Parklawn Drive, Rockville, Md. 20852 USA, on Nov. 1, 1995, under accession numbers 69943 and 69942, respectively.

Phage DNA was prepared from clones 14B1, 9F1, and 20A1. The 14B1 and 20A1 phage DNA were digested with Pst I to isolate the 1.2 kb and 1.5 kb fragments, respectively, that hybridized to the mouse neuroD probe. The 9F1 phage DNA was digested with Eco RI and SacI to obtain an approximately 2.2 kb fragment that hybridizes with the mouse neuroD probe. The fragments were each subcloned into plasmid BLUESCRIPT SK (Stratagene) that had been linearized with the appropriate restriction enzyme(s). The fragments were sequenced using Sequenase Version 2.0 (US Biochemical) and the following primers: the universal primer M13-21, the T7 primer, and the T3 primer. Sequence analysis of clones 9F1 (SEQ ID NOS:8 and 9), and 14B 1 (SEQ ID NOS:10 and 11) showed a high similarity between the mouse and human coding sequences at both the amino

acid and nucleotide level. In addition, while clones 9F1 and 14B1 shared 100% identity in the HLH region at the amino acid level (i.e., residues 117–156 in SEQ ID NO:9 and residues 137–176 in SEQ ID NO:11), they diverged in the amino-terminal of the bHLH. This finding strongly suggests that 14B1 is a member of the neuroD family of genes. Sequence analysis demonstrates that clone 9F1 has a high degree of homology throughout the sequence region that spans the translation start site to the end of the bHLH region. The 9F1 clone has 100% identity to mouse NeuroD in the HLH region (i.e., residues 117–156 in SEQ ID NO:9 and residues 117–156 in SEQ ID NO:2), and an overall identity of 94%. The 14B1 clone also has 100% identity to the HLH region (i.e., residues 137–176 in SEQ ID NO:11 and residues 117–156 in SEQ ID NO:2), but only 40% identity to 9F1 and 39% identity to mouse NeuroD in the amino-terminal region. This demonstrates that 9F1 is the human homolog of mouse neuroD, whereas the strong conservation of the NeuroD HLH identifies 14B1 as another member of the neuroD HLH subfamily. Human clone 9F1 (represented by SEQ ID NOS:8 and 9) is referred to as human neuroD. Human clone 14B 1 is referred to as neuroD2 (SEQ ID NOS:10 and 11, and human clone 20A1 is referred to as neuroD3 (SEQ ID NOS:12 and 13).

An 800 bp Hind III-Eag I fragment from the neuroD2 sequences from clone 14B1 was random primed with ^{32}P. This probe was used to screen a 16-day mouse embryo cDNA library essentially as described previously. Filters were prehybridized in FBI hybridization buffer (see above) at 50° C. for 10 minutes. After prehybridization, denatured salmon sperm DNA was added to a final concentration of 10 μg/ml; denatured probe was added to a final concentration of one million cpm/ml. The filter was hybridized at 50° C. overnight. After incubation, excess probe was removed, and the filter was washed first in 2× SSC, 0.1% SDS for 30 minutes at 60° C. One clone, designated 1.1.1, contained 1.46 kb of murine neuroD2 cDNA as an Eco RI-Hind IIII insert. The nucleotide sequence and deduced amino acid sequences are shown in SEQ ID NOS:16 and 17, respectively. A comparison between the human genomic sequence and the mouse cDNA sequence demonstrate that there were no introns in the human neuroD2 coding region.

In a similar manner, a random-primed 1.1 kb Pst I fragment from the human neuroD3 cDNA present in the 20A1 clone is prepared. The probe is used to screen a mouse embryo and newborn mouse brain libraries. Hybridization and wash conditions are as described above. Positive clones are analyzed by restriction and sequence analysis, and a full length clone is obtained. The mouse neuroD3 cDNA is used to prepare a probe for Northern analysis to study expression patterns in embryonic through adult mice.

Using a random-primed antisense probe to the mouse neuroD2 (Boehringer Mannheim) the expression pattern was determined using Northern analysis. Filters containing murine RNA from the brain and spinal cords of embryonic through adult mice were probed at high stringency and washed in 0.1× SSC, 0.1% SDS at 65° C. Northern analysis showed neuroD2 expression in the brain and spinal cords of mice from embryonic day 12.5 through adult.

EXAMPLE 12

Chromosome mapping of human neuroD clones

FISH karyotyping was performed on fixed metaphase spreads of the microcell hybrids essentially as described (Trask et al., *Am. J. Hum. Genet.* 48: 1–15, 1991; and Brandriff et al., *Genomics* 10: 75–82, 1991; which are incorporated by reference herein in their entirety). neuroD sequences were detected using the 9F1 or 20A1 phage DNA as probes labeled using digoxigenin dUTP (Boehringer Mannheim) according to the manufacturer's instructions. Phage DNA was biotinylated by random priming (Gibco/BRL BioNick Kit) and hybridized in situ to denatured metaphase chromosome spreads for 24–48 hours. Probes were detected with rhodamine-conjugated antibodies to digoxigenin, and chromosomes were counterstained with DAPI (Sigma). Signals were viewed through a fluorescence microscope and photographs were taken with color slide film. FISH analysis indicated clone 9F1 maps to human chromosome 2q, and clone 20A1 maps to human chromosome 5.

Chromosome mapping was also carried out on a human/rodent somatic cell hybrid panel (National Institute of General Medical Sciences, Camden, N.J.). This panel consists of DNA isolated from 24 human/rodent somatic cell hybrids retaining one human chromosome. For one set of experiments, the panel of DNA's were digested with Eco RI and electrophoresed on an agarose gel. The DNA was transferred to Hybond-N membranes (Amersham). A random primed (Boehringer Mannheim) 4 kb Eco RI-Sac I fragment of clone 9F1 was prepared. The filter was prehybridized in 10 ml of FBI hybridization buffer (see above) at 65° C. for 10 minutes. After prehybridization, denatured salmon sperm DNA was added to a final concentration of 10 μg/ml; denatured probe was added to a final concentration of one million cpm/ml. The filter was hybridized at 65° C. for a period of 8 hours to overnight. After incubation, excess probe was removed, and the filter was washed first in 2× SSC, 0.1% SDS for 30 minutes at 65° C. The first wash was followed by a final wash in 0.1× SSC, 0.1% SDS for 30 minutes at 65° C. An autoradiograph of the filter was prepared. Autoradiographs confirmed the FISH mapping results.

In the second experiment, the panel was digested with Pst I, electrophoresed and transferred essentially as described above. A random-primed (Boehringer Mannheim) 1.6 kb Psi I fragment of clone 20A1 was prepared. The membrane was prehybridized, hybridized with the 20A1 probe and washed as described above. Autoradiographs of the Southern showed that 20A1 mapped to human chromosome 5 and confirmed the FISH mapping results. After autoradiography, the 20A1-probed membrane was stripped by a wash in 0.5M NaOH, 1.5M NaCl. The membrane was neutralized in 0.5M Tris-HCl (pH 7.4), 1.5M NaCl. The filter was washed in 0.1× SSC before prehybridization. A random-primed (Boehringer Mannheim) 1.2 kb Pst I fragment of clone 14B1 was prepared. The washed membrane was prehybridized and hybridized with the 14B1 probe as described above. After washing under the previously described conditions, the membrane was autoradiographed. Autoradiographs demonstrated that clone 14B1 mapped to chromosome 17.

EXAMPLE 13

Human neuroD complementary DNA

To obtain a human neuroD cDNA, one million plaque forming units (pfu) were plated onto twenty LB+10 mM $MgSO_4$ (150 mm) plates using the Stratagene human cDNA library in Lambda ZAP II in the bacterial strain XL-1 Blue (Stratagene). Plating and membrane lifts were performed using standard methods, as described in Example 11. After UV cross-linking, the membranes were pre-hybridized in an aqueous hybridization solution (1% bovine serum albumin, 1 mM EDTA, 0.5M Na_2HPO_4 (pH 7.4), 7% SDS) at 50° C. for two hours.

The mouse neuroD cDNA insert was prepared by digesting the pKS+m7a RX plasmid with Eco RI and Xho I, and isolating the fragment containing the cDNA by electroelution. A probe was made with the cDNA containing fragment by random primed synthesis with random hexanucleotides, dGTP, dATP, dTTP, alpha-^{32}P-labeled dCTP, and Klenow in a buffered solution (25 mM Tris (pH6.9), 50 mM KCl, 5 mM $MgCl_2$, 1 mM DTT). The probe was purified from the unincorporated nucleotides on a G-50 sepharose column. The purified probe was heat denatured at 90° C. for 3 minutes.

After prehybridization, the denatured probe was added to the membranes in hybridization solution. The membranes were hybridized for 24 hours at 50° C. Excess probe was removed from the membranes, and the membranes were washed in 0.1× SSC, 0.1% SDS for 20 minutes at 50° C. The wash solution was changed five times. The membranes were blotted dry and covered with plastic film before being subjected to autoradiography. Autoradiography of the filters identified 68 positive clones. The clones are plaque-purified and rescreened to obtain 40 pure, positive clones. The positive clones were screened with a random-primed Pst I fragment from clone 9F1 (human neuroD). Twelve positive clones that hybridized with the human neuroD genomic probe were isolated.

The plasmid vector containing cDNA insert was excised in vivo from the lambda phage clone according to the Stratagene methodology. Briefly, eluted phage and XL-1 Blue cells (200 microliters of OD 600=1) were mixed with R408 helper phage provided by Stratagene for 15 minutes at 37° C. Five milliliters of rich bacterial growth media (2 X YT, see Sambrook et at., ibid.) was added, and the cultures were incubated for 3 hours at 37° C. The tubes were heated at 70° C. for 20 minutes and spun for 5 minutes at 4,000×g. After centrifugation, 200 microliters of supernatant was added to the same volume of XL-1 Blue cells (OD=1), and the mixture was incubated for 15 minutes at 37° C, after which the bacterial cells were plated onto LB plates containing 50 μg/ml ampicillin. Each colony was picked and grown for sequencing template preparation. The clones were sequenced and compared to the human genomic sequence. A full length cDNA encoding human neuroD that was identical to the 9F1 neuroD genomic sequence was obtained and designated HC2A. The nucleotide and deduced amino acid sequences are shown in SEQ ID NOS:14 and 15 , respectively. Clone HC2A was deposited at the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Md. 20852 USA, on Nov. 1, 1995, under accession number 69944.

EXAMPLE 14

Construction of knock-out mice

Knock-out mice in which the murine neuroD coding sequence was replaced with the β-galactosidase gene and the neomycin resistance gene (neo) were generated i) to assess the consequences of eliminating the murine NeuroD protein during mouse development and ii) to permit examination of the expression pattern of neuroD in embryonic mice. Genomic neuroD sequences used for these knock-out mice were obtained from the 129/Sv mice so that the homologous recombination could take place in a congenic background in 129/Sv mouse embryonic stem cells. Several murine neuroD genomic clones were isolated from a genomic library prepared from 129/Sv mice (Zhuang et at., *Cell* 79: 875–884, 1994; which is incorporated herein by reference in its entirety) using the Bam HI-Not I neuroD cDNA containing fragment of pSK+1-83 (Example 2) as a random-primed probe essentially as described in Example 11. Plasmid pPNT (Tybulewicz et at., *Cell* 65: 1153–1163, 1991; which is incorporated herein by reference in its entirety) containing the neomycin resistance gene (neo; a positive selection marker) and the *Herpes simplex* virus thymidine kinase gene (hsv-tk, a negative selection marker) under the control of the PGK promoter provided the vector backbone for the targeting construct. A 1.4 kb 5' murine neuroD genomic fragment together with the 3 kb cytoplasmic β-galactosidase gene were inserted between the Eco RI and Xba I sites of the pPNT vector, and an 8 kb fragment containing the genomic 3' untranslated sequence of neuroD was inserted into the vector backbone between into the Xho I and Not I sites.

To prepare an Eco RI-Xba I fragment containing neuroD promoter sequences joined to the β-galactosidase gene, a 1.4 kb Eco RI(vector-derived)-Asp718 fragment containing the 5' untranslated murine neuroD genomic sequence was ligated to a Hind III-Xba I fragment containing the cytoplasmic β-galactosidase gene such that the Asp 718 and Hind III sites were destroyed. The resulting approximately 4.4 kb Eco RI-Xba I fragment, containing the 5' neuroD genomic sequence (including the neuroD promoter) and the β-galactosidase gene in the same transcriptional orientation, was inserted into Eco RI-Xba I linearized pPNT to yield the plasmid pPNT/5'+β-gal. A neuroD fragment containing 3' untranslated DNA was obtained from a murine neuroD genomic clone that had been digested with Spe I and Not I(vector-derived) to yield an 8 kb fragment. To obtain a 5' Xho I site, the 8 kb fragment was inserted into Spe I-Not I linearized pBluescriptSK+(Stratagene), and the resulting plasmid digested with Xho I and Not I to obtain the 8 kb neuroD 3' genomic fragment. The Xho I-Not I fragment was inserted into Xho I-Not I linearized pPNT/5'+β-gal to yield the neuroD targeting vector. The final construct contained the 5' neuroD fragment, the β-galactosidase gene, and the 3' genomic neuroD fragment in the same orientation, and the hsv-tk and neomycin resistance genes in the opposite orientation.

The targeting construct was transfected by electroporation into mouse embryonic stem (ES) cells. A 129/Sv derived ES cell line, AK-7 described by Zhuang et at. (ibid.) was used for electroporation. These ES cells were routinely cultured on mitomycin C-treated (Sigma) SNL 76/7 cells (feeder cells) as described by McMahon and Bradley (*Cell* 62: 1073–1085, 1990; which is incorporated herein by reference in its entirety) in culture medium containing high glucose DMEM supplemented with 15% fetal bovine serum (Hyclone) and 0. 1 μM β-mercaptoethanol. To prepare the targeting construct for transfection, 25 μg of the targeting construct was linearized by digestion with Not I, phenol-chloroform extracted, and ethanol precipitated. The linearized vector was then electroporated into 1–2×10^7 AK-7 (ES) cells. The electroporated cells were seeded onto three 10-cm plates, with one plate receiving 50% of the electroporated cells and the remaining two plates each receiving 25% of the electroporated cells. After 24 hours, G418 was added to each of the plates to a final concentration of 150 μ/ml. After an additional 24 hours, gancyclovir was added to a final concentration of 0.2 μM to the 50% plate and one of the 25% plates. The third plate containing 25% of the electroporated cells was subjected to only G418 selection to assess the efficiency of gancyclovir selection. The culture medium for each plate was changed every day for the first few days, and then changed as needed after selection had occurred. After 10 days of selection, a portion of each colony was picked microscopically with a drawn micropipette, and was directly

analyzed by PCR as described by Joyner et al. (*Nature* 338: 153–156, 1989; which is incorporated herein by reference in its entirety). Briefly, PCR amplification was performed as described (Kogan et at., *New England J. Med.* 317: 985–990, 1987; which is incorporated herein by reference in its entirety) using 40 cycles of 93° C. for 30 seconds, 57° C. for 30 seconds, and 65° C. for 3 minutes. To detect the wild-type allele, primers JL34 and JL36 (SEQ ID NOS:18 and 19, respectively) were used in the PCR reaction, to detect the mutant neuroD allele, primers JL34 and JL40 (SEQ ID NOS:18 and 20, respectively) were used in the PCR reaction. Positive colonies, identified by PCR, were subcloned into 4-well plates, expanded into 60 turn plates and frozen into 2–3 ampules.

Among the clones that were selected for both G418-resistance (positive selection for neo gene expression) and gancyclovir-resistance (negative selection for hsv-tk gene expression), 10% of the population contained correctly targeted integration of the vector into the murine neuroD locus (an overall 10% targeting frequency) The negative selection provided 4–3 fold enrichment for homologous recombination events.

To generate chimeric mice, each positive clone was thawed and passaged once on feeder cells. The transfected cells were trypsinized into single cells, and blastocysts obtained from C57BL/6J mice were injected with approximately 15 cells. The injected blastocysts were then implanted into pseudopregnant mice (C57BL/6J×CBA). Four male chimeras arose from the injected blastocysts (AK-71, AK-72, AK-74 and AK-75). The male chimeras AK-71 and AK-72 gave germ-line transmission at a high rate as determined by the frequency of agouti coat color transmission to their offspring (F1) in a cross with C57BL/6J female mice. Since 50% of the agouti coat color offspring (F1) should represent heterozygous mutants, their genotypes were determined by Southern blot analysis. Briefly, genomic DNA prepared from tail biopsies was digested with Eco RI and probed with the 1.4 kb 5' genomic sequence used to make the targeting construct. This probe detects a 4 kb Eco RI fragment from the wild-type allele and a 6.3 kb Eco RI fragment from the mutant allele. Therefore, a Southern analysis would show a single 4 kb band for a wild-type mouse, 4 kb and 6.3 kb fragments for a heterozygous mouse, and a single 6.3 kb band for a homozygous mutant mouse. The resulting offspring (F1) heterozygous (±) mice, were mated with sibling heterozygous mice to give rise to the homozygous (–/–) mutant mice.

To study neuroD expression patterns in embryonic mice, chimeric mice or F1 heterozygous progeny from the chimera x C57B/6J mating were crossed with C57B/6J. Litters resulting from these crosses were harvested from pregnant females and stained for β-galactosidase activity. The embryos were dissected away from all the extra-embryonic tissue and the yolk sac was reserved for DNA analysis. The embryos were fixed for one hour in a Fix solution (0.1M phosphate buffer containing 0.2% glutaraldehyde, 2% formaldehyde, 5 mM EGTA)pH 7.3), 2 mM $MgCl_2$). The fixing solution was removed by three thirty-minute rinses with rinse solution (0.1M phosphate buffer (pH 7.3) containing 2 mM $MgCl_2$, 0.1% sodium deoxycholate, 0.2% NP-40). The fixed embryos were stained overnight in the dark in rinse solution containing 1 mg/ml X-gal, 5 mM sodium ferricyanide, 5 mM sodium ferrocyanide. After staining, the embryos were rinsed with PBS and stored in the Fix solution before preparation for examination. Examination of stained tissue from fetal and postnatal mice heterozygous for the mutation confirmed neuroD expression pattern in neuronal cells demonstrated by in situ hybridization (Example 4) and also demonstrated neuroD expression in the pancreas and gastrointestinal tract.

Blood glucose levels were detected using PRECISION QID blood glucose test strips and a PRECISION QID blood glucose sensor (Medisens Inc., Waltham, Mass.) according to the manufacturer's instruction. A tissue sample was taken for DNA analysis and the pups were fixed for further histological examination. Blood glucose levels in mice homozygous for the mutation (neuroD) had blood glucose levels between 2 and 3 times higher than the blood glucose level of wild-type mice. Heterozygous mutants exhibited similar blood glucose levels as wild-type mice. Mice that were homozygous for the mutation (lacking neuroD) had diabetes as demonstrated by high blood glucose levels and died by day four; some homozygous mice died at birth.

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modification may be made without deviating from the spirit and scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 20

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2089 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Mus musculus

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 229..1302

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ACTACGCAGC ACCGAGGTAC AGACACGCCA GCATGAAGCA CTGCGTTTAA CTTTTCCTGG      60

AGGCATCCAT TTTGCAGTGG ACTCCTGTGT ATTTCTATTT GTGTGCATTT CTGTAGGATT     120

AGGGAGAGGG AGCTGAAGGC TTATCCAGCT TTTAAATATA GCGGGTGGAT TTCCCCCCCT     180

TTCTTCTTCT GCTTGCCTCT CTCCCTGTTC AATACAGGAA GTGGAAAC ATG ACC AAA      237
                                                    Met Thr Lys
                                                     1

TCA TAC AGC GAG AGC GGG CTG ATG GGC GAG CCT CAG CCC CAA GGT CCC      285
Ser Tyr Ser Glu Ser Gly Leu Met Gly Glu Pro Gln Pro Gln Gly Pro
  5                  10                  15

CCA AGC TGG ACA GAT GAG TGT CTC AGT TCT CAG GAC GAG GAA CAC GAG      333
Pro Ser Trp Thr Asp Glu Cys Leu Ser Ser Gln Asp Glu Glu His Glu
 20                  25                  30                  35

GCA GAC AAG AAA GAG GAC GAG CTT GAA GCC ATG AAT GCA GAG GAG GAC      381
Ala Asp Lys Lys Glu Asp Glu Leu Glu Ala Met Asn Ala Glu Glu Asp
                 40                  45                  50

TCT CTG AGA AAC GGG GGA GAG GAG GAG GAG GAA GAT GAG GAT CTA GAG      429
Ser Leu Arg Asn Gly Gly Glu Glu Glu Glu Glu Asp Glu Asp Leu Glu
             55                  60                  65

GAA GAG GAG GAA GAA GAA GAG GAG GAG GAG GAT CAA AAG CCC AAG AGA      477
Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Asp Gln Lys Pro Lys Arg
         70                  75                  80

CGG GGT CCC AAA AAG AAA AAG ATG ACC AAG GCG CGC CTA GAA CGT TTT      525
Arg Gly Pro Lys Lys Lys Lys Met Thr Lys Ala Arg Leu Glu Arg Phe
     85                  90                  95

AAA TTA AGG CGC ATG AAG GCC AAC GCC CGC GAG CGG AAC CGC ATG CAC      573
Lys Leu Arg Arg Met Lys Ala Asn Ala Arg Glu Arg Asn Arg Met His
100                 105                 110                 115

GGG CTG AAC GCG GCG CTG GAC AAC CTG CGC AAG GTG GTA CCT TGC TAC      621
Gly Leu Asn Ala Ala Leu Asp Asn Leu Arg Lys Val Val Pro Cys Tyr
                120                 125                 130

TCC AAG ACC CAG AAA CTG TCT AAA ATA GAG ACA CTG CGC TTG GCC AAG      669
Ser Lys Thr Gln Lys Leu Ser Lys Ile Glu Thr Leu Arg Leu Ala Lys
            135                 140                 145

AAC TAC ATC TGG GCT CTG TCA GAG ATC CTG CGC TCA GGC AAA AGC CCT      717
Asn Tyr Ile Trp Ala Leu Ser Glu Ile Leu Arg Ser Gly Lys Ser Pro
        150                 155                 160

GAT CTG GTC TCC TTC GTA CAG ACG CTC TGC AAA GGT TTG TCC CAG CCC      765
Asp Leu Val Ser Phe Val Gln Thr Leu Cys Lys Gly Leu Ser Gln Pro
    165                 170                 175

ACT ACC AAT TTG GTC GCC GGC TGC CTG CAG CTC AAC CCT CGG ACT TTC      813
Thr Thr Asn Leu Val Ala Gly Cys Leu Gln Leu Asn Pro Arg Thr Phe
180                 185                 190                 195

TTG CCT GAG CAG AAC CCG GAC ATG CCC CCG CAT CTG CCA ACC GCC AGC      861
Leu Pro Glu Gln Asn Pro Asp Met Pro Pro His Leu Pro Thr Ala Ser
                200                 205                 210

GCT TCC TTC CCG GTG CAT CCC TAC TCC TAC CAG TCC CCT GGA CTG CCC      909
Ala Ser Phe Pro Val His Pro Tyr Ser Tyr Gln Ser Pro Gly Leu Pro
            215                 220                 225

AGC CCG CCC TAC GGC ACC ATG GAC AGC TCC CAC GTC TTC CAC GTC AAG      957
Ser Pro Pro Tyr Gly Thr Met Asp Ser Ser His Val Phe His Val Lys
        230                 235                 240

CCG CCG CCA CAC GCC TAC AGC GCA GCT CTG GAG CCC TTC TTT GAA AGC     1005
Pro Pro Pro His Ala Tyr Ser Ala Ala Leu Glu Pro Phe Phe Glu Ser
    245                 250                 255

CCC CTA ACT GAC TGC ACC AGC CCT TCC TTT GAC GGA CCC CTC AGC CCG     1053
Pro Leu Thr Asp Cys Thr Ser Pro Ser Phe Asp Gly Pro Leu Ser Pro
260                 265                 270                 275
```

-continued

```
CCG CTC AGC ATC AAT GGC AAC TTC TCT TTC AAA CAC GAA CCA TCC GCC      1101
Pro Leu Ser Ile Asn Gly Asn Phe Ser Phe Lys His Glu Pro Ser Ala
                280                 285                 290

GAG TTT GAA AAA AAT TAT GCC TTT ACC ATG CAC TAC CCT GCA GCG ACG      1149
Glu Phe Glu Lys Asn Tyr Ala Phe Thr Met His Tyr Pro Ala Ala Thr
            295                 300                 305

CTG GCA GGG CCC CAA AGC CAC GGA TCA ATC TTC TCT TCC GGT GCC GCT      1197
Leu Ala Gly Pro Gln Ser His Gly Ser Ile Phe Ser Ser Gly Ala Ala
        310                 315                 320

GCC CCT CGC TGC GAG ATC CCC ATA GAC AAC ATT ATG TCT TTC GAT AGC      1245
Ala Pro Arg Cys Glu Ile Pro Ile Asp Asn Ile Met Ser Phe Asp Ser
    325                 330                 335

CAT TCG CAT CAT GAG CGA GTC ATG AGT GCC CAG CTT AAT GCC ATC TTT      1293
His Ser His His Glu Arg Val Met Ser Ala Gln Leu Asn Ala Ile Phe
340                 345                 350                 355

CAC GAT TAGAGGGCAC GTCAGTTTCA CTATTCCCGG GAAACGAATC CACTGTGCGT       1349
His Asp

ACAGTGACTG TCCTGTTTAC AGAAGGCAGC CCTTTTGCTA AGATTGCTGC AAAGTGCAAA   1409

TACTCAAAGC TTCAAGTGAT ATATGTATTT ATTGTCGTTA CTGCCTTTGG AAGAAACAGG   1469

GGATCAAAGT TCCTGTTCAC CTTATGTATT GTTTTCTATA GCTCTTCTAT TTTAAAAATA   1529

ATAATACAGT AAAGTAAAAA AGAAAATGTG TACCACGAAT TTCGTGTAGC TGTATTCAGA   1589

TCGTATTAAT TATCTGATCG GGATAAAAAA AATCACAAGC AATAATTAGG ATCTATGCAA   1649

TTTTTAAACT AGTAATGGGC CAATTAAAAT ATATATAAAT ATATATTTTT CAACCAGCAT   1709

TTTACTACCT GTGACCTTTC CCATGCTGAA TTATTTTGTT GTGATTTTGT ACAGAATTTT   1769

TAATGACTTT TTATAACGTG GATTTCCTAT TTTAAAACCA TGCAGCTTCA TCAATTTTTA   1829

TACATATCAG AAAAGTAGAA TTATATCTAA TTTATACAAA ATAATTTAAC TAATTTAAAC   1889

CAGCAGAAAA GTGCTTAGAA AGTTATTGCG TTGCCTTAGC ACTTCTTTCT TCTCTAATTG   1949

TAAAAAAGAA AAAAAAAAAA AAAAAACTCG AGGGGGGGCC CGGTACCCAG CTTTTGTTCC   2009

CTTTAGTGAG GGTTAATTGC GCGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA   2069

ATTGTTATCC GCTCACAATT                                               2089
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 357 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Thr Lys Ser Tyr Ser Glu Ser Gly Leu Met Gly Glu Pro Glu Pro
 1               5                  10                  15

Gln Gly Pro Pro Ser Trp Thr Asp Glu Cys Leu Ser Ser Gln Asp Glu
            20                  25                  30

Glu His Glu Ala Asp Lys Lys Glu Asp Glu Leu Glu Ala Met Asn Ala
        35                  40                  45

Glu Glu Asp Ser Leu Arg Asn Gly Gly Glu Glu Glu Glu Glu Asp Glu
    50                  55                  60

Asp Leu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Asp Gln Lys
65                  70                  75                  80

Pro Lys Arg Arg Gly Pro Lys Lys Lys Lys Met Thr Lys Ala Arg Leu
                85                  90                  95
```

-continued

```
Glu Arg Phe Lys Leu Arg Arg Met Lys Ala Asn Ala Arg Glu Arg Asn
                100                 105                 110

Arg Met His Gly Leu Asn Ala Ala Leu Asp Asn Leu Arg Lys Val Val
            115                 120                 125

Pro Cys Tyr Ser Lys Thr Gln Lys Leu Ser Lys Ile Glu Thr Leu Arg
        130                 135                 140

Leu Ala Lys Asn Tyr Ile Trp Ala Leu Ser Glu Ile Leu Arg Ser Gly
145                 150                 155                 160

Lys Ser Pro Asp Leu Val Ser Phe Val Glu Thr Leu Cys Lys Gly Leu
                165                 170                 175

Ser Gln Pro Thr Thr Asn Leu Val Ala Gly Cys Leu Gln Leu Asn Pro
            180                 185                 190

Arg Thr Phe Leu Pro Glu Gln Asn Pro Asp Met Pro Pro His Leu Pro
        195                 200                 205

Thr Ala Ser Ala Ser Phe Pro Val His Pro Tyr Ser Tyr Gln Ser Pro
    210                 215                 220

Gly Leu Pro Ser Pro Pro Tyr Gly Thr Met Asp Ser Ser His Val Phe
225                 230                 235                 240

His Val Lys Pro Pro Pro His Ala Tyr Ser Ala Ala Leu Glu Pro Phe
                245                 250                 255

Phe Glu Ser Pro Leu Thr Asp Cys Thr Ser Pro Ser Phe Asp Gly Pro
            260                 265                 270

Leu Ser Pro Pro Leu Ser Ile Asn Gly Asn Phe Ser Phe Lys His Glu
        275                 280                 285

Pro Ser Ala Glu Phe Glu Lys Asn Tyr Ala Phe Thr Met His Tyr Pro
    290                 295                 300

Ala Ala Thr Leu Ala Gly Pro Gln Ser His Gly Ser Ile Phe Ser Ser
305                 310                 315                 320

Gly Ala Ala Ala Pro Arg Cys Glu Ile Pro Ile Asp Asn Ile Met Ser
                325                 330                 335

Phe Asp Ser His Ser His His Glu Arg Val Met Ser Ala Gln Leu Asn
            340                 345                 350

Ala Ile Phe His Asp
        355
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1275 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Xenopus laevis

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 25..1083

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATTTCCTTTC TCCAGATCTA AAAA ATG ACC AAA TCG TAT GGA GAG AAT GGG      51
                           Met Thr Lys Ser Tyr Gly Glu Asn Gly
                             1               5

CTG ATC CTG GCC GAG ACT CCG GGC TGC AGA GGA TGG GTG GAC GAA TGC      99
Leu Ile Leu Ala Glu Thr Pro Gly Cys Arg Gly Trp Val Asp Glu Cys
 10                  15                  20                  25

CTG AGT TCT CAG GAT GAA AAC GAT CTG GAG AAA AAG GAG GGA GAG TTG     147
```

-continued

```
Leu Ser Ser Gln Asp Glu Asn Asp Leu Glu Lys Lys Glu Gly Glu Leu
                30                  35                  40

ATG AAA GAA GAC GAT GAA GAC TCA CTG AAT CAT CAC AAT GGA GAG GAG      195
Met Lys Glu Asp Asp Glu Asp Ser Leu Asn His His Asn Gly Glu Glu
            45                  50                  55

AAC GAG GAA GAG GAT GAA GGG GAT GAG GAG GAG GAG GAC GAT GAA GAT      243
Asn Glu Glu Glu Asp Glu Gly Asp Glu Glu Glu Glu Asp Asp Glu Asp
        60                  65                  70

GAT GAT GAG GAT GAC GAC CAG AAA CCC AAA AGG CGA GGA CCG AAA AAG      291
Asp Asp Glu Asp Asp Asp Gln Lys Pro Lys Arg Arg Gly Pro Lys Lys
    75                  80                  85

AAA AAA ATG ACG AAA GCC CGG GTG GAG CGA TTT AAA GTG AGA CGC ATG      339
Lys Lys Met Thr Lys Ala Arg Val Glu Arg Phe Lys Val Arg Arg Met
 90                  95                 100                 105

AAG GCA AAC GCC AGG GAG AGG AAT CGC ATG CAC GGA CTC AAC GAT GCC      387
Lys Ala Asn Ala Arg Glu Arg Asn Arg Met His Gly Leu Asn Asp Ala
                110                 115                 120

CTG GAC AGT CTG CGC AAA GTT GTG CCC TGC TAC TCC AAA ACA CAA AAG      435
Leu Asp Ser Leu Arg Lys Val Val Pro Cys Tyr Ser Lys Thr Gln Lys
            125                 130                 135

TTG TCT AAG ATT GAA ACT CTG CGC CTG GCT AAG AAC TAC ATC TGG GCT      483
Leu Ser Lys Ile Glu Thr Leu Arg Leu Ala Lys Asn Tyr Ile Trp Ala
        140                 145                 150

CTT TCT GAG ATT TTA AGG TCC GGC AAA AGC CCA GAC CTG GTG TCC TTT      531
Leu Ser Glu Ile Leu Arg Ser Gly Lys Ser Pro Asp Leu Val Ser Phe
    155                 160                 165

GTA CAA ACT CTC TGC AAA GGT TTG TCG CAG CCC ACC ACC AAT CTA GTA      579
Val Gln Thr Leu Cys Lys Gly Leu Ser Gln Pro Thr Thr Asn Leu Val
170                 175                 180                 185

GCG GGG TGT CTG CAG CTG AAC CCC AGA ACT TTC CTT CCT GAG CAG AGT      627
Ala Gly Cys Leu Gln Leu Asn Pro Arg Thr Phe Leu Pro Glu Gln Ser
                190                 195                 200

CAG GAC ATC CAG TCG CAC ATG CAA ACA GCG AGC TCT TCC TTC CCT CTG      675
Gln Asp Ile Gln Ser His Met Gln Thr Ala Ser Ser Ser Phe Pro Leu
            205                 210                 215

CAG GGC TAT CCC TAT CAG TCC CCT GGT CTT CCC AGT CCC CCC TAT GGT      723
Gln Gly Tyr Pro Tyr Gln Ser Pro Gly Leu Pro Ser Pro Pro Tyr Gly
        220                 225                 230

ACC ATG GAC AGC TCC CAT GTA TTC CAC GTC AAG CCT CAC TCC TAT GGG      771
Thr Met Asp Ser Ser His Val Phe His Val Lys Pro His Ser Tyr Gly
    235                 240                 245

GCG GCC CTG GAG CCT TTC TTT GAC AGC AGC ACC GTC ACT GAG TGT ACC      819
Ala Ala Leu Glu Pro Phe Phe Asp Ser Ser Thr Val Thr Glu Cys Thr
250                 255                 260                 265

AGC CCG TCA TTC GAT GGT CCC CTG AGC CCA CCC CTT AGT GTT AAT GGG      867
Ser Pro Ser Phe Asp Gly Pro Leu Ser Pro Pro Leu Ser Val Asn Gly
                270                 275                 280

AAC TTT ACT TTT AAA CAC GAG CAT TCG GAG TAT GAT AAA AAT TAC ACG      915
Asn Phe Thr Phe Lys His Glu His Ser Glu Tyr Asp Lys Asn Tyr Thr
            285                 290                 295

TTC ACT ATG CAC TAT CCT GCA GCC ACT ATA TCC CAG GGC CAC GGA CCA      963
Phe Thr Met His Tyr Pro Ala Ala Thr Ile Ser Gln Gly His Gly Pro
        300                 305                 310

TTG TTC TCC ACG GGG GGA CCA CGC TGT GAA ATC CCA ATA GAC ACC ATC     1011
Leu Phe Ser Thr Gly Gly Pro Arg Cys Glu Ile Pro Ile Asp Thr Ile
    315                 320                 325

ATG TCC TAT GAC GGT CAC TCC CAC CAT GAA AGA GTC ATG AGT GCC CAG     1059
Met Ser Tyr Asp Gly His Ser His His Glu Arg Val Met Ser Ala Gln
330                 335                 340                 345

CTA AAT GCC ATC TTT CAT GAT TAACCCTTGG AAGATCAAAA CAACTGACTG        1110
```

-continued

```
Leu Asn Ala Ile Phe His Asp
                350

TGCATTGCCA GGACTGTCTT GTTTACCAAG GGCAGACACG TGGGTAGTAA AAGTGCAAAT   1170

GCCCCACTCT GGGGCTGTAA CAAACTTGAT CTTGTCCTGC CTTTAGATAT GGGGAAACCT   1230

AATGTATTAA TTCCCACCTC CTTCCAATCG ACACTCCTTT AAATT                   1275
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 352 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Thr Lys Ser Tyr Gly Glu Asn Gly Leu Ile Leu Ala Glu Thr Pro
 1               5                  10                  15

Gly Cys Arg Gly Trp Val Asp Glu Cys Leu Ser Ser Glu Asp Glu Asn
            20                  25                  30

Asp Leu Glu Lys Lys Glu Gly Glu Leu Met Lys Glu Asp Asp Glu Asp
        35                  40                  45

Ser Leu Asn His His Asn Gly Glu Glu Asn Glu Glu Glu Asp Glu Gly
    50                  55                  60

Asp Glu Glu Glu Glu Asp Asp Glu Asp Asp Asp Glu Asp Asp Asp Gln
65                  70                  75                  80

Lys Pro Lys Arg Arg Gly Pro Lys Lys Lys Lys Met Thr Lys Ala Arg
                85                  90                  95

Val Glu Arg Phe Lys Val Arg Arg Met Lys Ala Asn Ala Arg Glu Arg
            100                 105                 110

Asn Arg Met His Gly Leu Asn Asp Ala Leu Asp Ser Leu Arg Lys Val
        115                 120                 125

Val Pro Cys Tyr Ser Lys Thr Gln Lys Leu Ser Lys Ile Glu Thr Leu
    130                 135                 140

Arg Leu Ala Lys Asn Tyr Ile Trp Ala Leu Ser Glu Ile Leu Arg Ser
145                 150                 155                 160

Gly Lys Ser Pro Asp Leu Val Ser Phe Val Gln Thr Leu Cys Lys Gly
                165                 170                 175

Leu Ser Gln Pro Thr Thr Asn Leu Val Ala Gly Cys Leu Gln Leu Asn
            180                 185                 190

Pro Arg Thr Phe Leu Pro Glu Gln Ser Glu Asp Ile Gln Ser His Met
        195                 200                 205

Gln Thr Ala Ser Ser Ser Phe Pro Leu Gln Gly Tyr Pro Tyr Gln Ser
    210                 215                 220

Pro Gly Leu Pro Ser Pro Pro Tyr Gly Thr Met Asp Ser Ser His Val
225                 230                 235                 240

Phe His Val Lys Pro His Ser Tyr Gly Ala Ala Leu Glu Pro Phe Phe
                245                 250                 255

Asp Ser Ser Thr Val Thr Glu Cys Thr Ser Pro Ser Phe Asp Gly Pro
            260                 265                 270

Leu Ser Pro Pro Leu Ser Val Asn Gly Asn Phe Thr Phe Lys His Glu
        275                 280                 285

His Ser Glu Tyr Asp Lys Asn Tyr Thr Phe Thr Met His Tyr Pro Ala
    290                 295                 300

Ala Thr Ile Ser Gln Gly His Gly Pro Leu Phe Ser Thr Gly Gly Pro
305                 310                 315                 320
```

-continued

```
Arg Cys Glu Ile Pro Ile Asp Thr Ile Met Ser Tyr Asp Gly His Ser
                325                 330                 335

His His Glu Arg Val Met Ser Ala Gln Leu Asn Ala Ile Phe His Asp
            340                 345                 350
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Asn Ala Arg Glu Arg Arg Arg
1               5
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Asn Glu Arg Glu Arg Asn Arg
1               5
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Asn Ala Arg Glu Arg
1               5
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 524 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: 9F1

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 57..524

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:8:

-continued

```
TTTTTCTGCT TTTCTTTCTG TTTGCCTCTC CCTTGTTGAA TGTAGGAAAT CGAAAC                56

ATG ACC AAA TCG TAC AGC GAG AGT GGG CTG ATG GGC GAG CCT CAG CCC              104
Met Thr Lys Ser Tyr Ser Glu Ser Gly Leu Met Gly Glu Pro Gln Pro
 1               5                  10                  15

CAA GGT CCT CCA AGC TGG ACA GAC GAG TGT CTC AGT TCT CAG GAC GAG              152
Gln Gly Pro Pro Ser Trp Thr Asp Glu Cys Leu Ser Ser Gln Asp Glu
            20                  25                  30

GAG CAC GAG GCA GAC AAG AAG GAG GAC GAC CTC GAA GCC ATG AAC GCA              200
Glu His Glu Ala Asp Lys Lys Glu Asp Asp Leu Glu Ala Met Asn Ala
        35                  40                  45

GAG GAG GAC TCA CTG AGG AAC GGG GGA GAG GAG GAG GAC GAA GAT GAG              248
Glu Glu Asp Ser Leu Arg Asn Gly Gly Glu Glu Glu Asp Glu Asp Glu
    50                  55                  60

GAC CTG GAA GAG GAG GAA GAA GAG GAA GAG GAG GAG GAT GAC GAT CAA AAG          296
Asp Leu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Asp Asp Asp Gln Lys
65                  70                  75                  80

CCC AAG AGA CGC GGC CCC AAA AAG AAG AAG ATG ACT AAG GCT CGC CTG              344
Pro Lys Arg Arg Gly Pro Lys Lys Lys Lys Met Thr Lys Ala Arg Leu
                85                  90                  95

GAG CGT TTT AAA TTG AGA CGC ATG AAG GCT AAC GCC CGG GAG CGG AAC              392
Glu Arg Phe Lys Leu Arg Arg Met Lys Ala Asn Ala Arg Glu Arg Asn
            100                 105                 110

CGC ATG CAC GGA CTG AAC GCG GCG CTA GAC AAC CTG CGC AAG GTG GTG              440
Arg Met His Gly Leu Asn Ala Ala Leu Asp Asn Leu Arg Lys Val Val
        115                 120                 125

CCT TGC TAT TCT AAG ACG CAG AAG CTG TCC AAA ATC GAG ACT CTG CGC              488
Pro Cys Tyr Ser Lys Thr Gln Lys Leu Ser Lys Ile Glu Thr Leu Arg
    130                 135                 140

TTG GCC AAG AAC TAC ATC TGG GCT CTG TCG GAG ATC                              524
Leu Ala Lys Asn Tyr Ile Trp Ala Leu Ser Glu Ile
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 156 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
Met Thr Lys Ser Tyr Ser Glu Ser Gly Leu Met Gly Glu Pro Gln Pro
 1               5                  10                  15

Gln Gly Pro Pro Ser Trp Thr Asp Glu Cys Leu Ser Ser Gln Asp Glu
            20                  25                  30

Glu His Glu Ala Asp Lys Lys Glu Asp Asp Leu Glu Ala Met Asn Ala
        35                  40                  45

Glu Glu Asp Ser Leu Arg Asn Gly Gly Glu Glu Glu Asp Glu Asp Glu
    50                  55                  60

Asp Leu Glu Glu Glu Glu Glu Glu Glu Glu Glu Asp Asp Asp Gln Lys
65                  70                  75                  80

Pro Lys Arg Arg Gly Pro Lys Lys Lys Lys Met Thr Lys Ala Arg Leu
                85                  90                  95

Glu Arg Phe Lys Leu Arg Arg Met Lys Ala Asn Ala Arg Glu Arg Asn
            100                 105                 110

Arg Met His Gly Leu Asn Ala Ala Leu Asp Asn Leu Arg Lys Val Val
        115                 120                 125

Pro Cys Tyr Ser Lys Thr Gln Lys Leu Ser Lys Ile Glu Thr Leu Arg
```

-continued

```
      130                 135                 140

Leu Ala Lys Asn Tyr Ile Trp Ala Leu Ser Glu Ile
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1352 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: 14B1

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 55..1194

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
CCCCTCACTT TGTGCTGTCT GTCTCCCCTT CCCGCCCGGG GNCCCTCAGG CACCATGCTG      60
ACCCGCCTGT TCAGCGAGCC CGGCCTTCTC TCGGACGTGC CCAAGTTCGC CAGCTGGGGC     120
GACGGCGAAG ACGACGAGCC GAGGAGCGAC AAGGGCGACG CGCCGCCACC GCCACCGCCT     180
GCGCCCGGGC CAGGGGCTCC GGGGCCAGCC CGGGCGGCCA AGCCAGTCCC TCTCCGTGGA     240
GAAGAGGGGA CGGAGGCCAC GTTGGCCGAG GTCAAGGAGG AAGGCGAGCT GGGGGGAGAG     300
GAGGAGGAGG AAGAGGAGGA GGAAGAAGGA CTGGACGAGG CGGAGGGCGA GCGGCCCAAG     360
AAGCGCGGGC CCAAGAAGCG CAAGATGACC AAGGCGCGCT TGGAGCGCTC CAAGCTTCGG     420
CGGCAGAAGG CGAACGCGCG GGAGCGCAAC CGCATGCACG ACCTGAACGC AGCCCTGGAC     480
AACCTGCGCA AGGTGGTGCC CTGCTACTCC AAGACGCAGA AGCTGTCCAA GATCGAGACG     540
CTGCGCCTAG CCAAGAACTA TATCTGGGCG CTCTCGGAGA TCCTGCGCTC CGGCAAGCGG     600
CCAGACCTAG TGTCCTACGT GCAGACTCTG TGCAAGGGTC TGTCGCAGCC CACCACCAAT     660
CTGGTGGCCG GCTGTCTGCA GCTCAACTCT CGCAACTTCC TCACGGAGCA AGGCCGCGAC     720
GGTGCGNNCC GCTTCCACGG CTCGGGCGGC CCGTTCGCCA TGCACCCCTA CCCGTACCCG     780
TGCTCGCGTG GCGGGCGGAC AGTGCCAGGC GCGGCGGCCT GGGCGGCGGC CGGCGCACGC     840
CTGCGGACCC ACGGCTACTG CGCCGCCTAC GAGACGCTGT ATGCGGCGGC AGGCGGTGGC     900
GGCGCGAGCC CGGACTACAA CAGCTCCGAG TACGAGGGCC CGCTCAGCCC CCCGCTCTGT     960
CTCAATGGCA ACTTCTCACT CAAGCAGGAC TCCTCGCCCG ACCACGAGAA AAGCTACCAC    1020
TACTCTATGC ACTACTCGGG CTGCCCNGGT TCGCGCCACG GNCACGGGCT AGTCTTCGGC    1080
TCGTCGGCTG TGCGCGGGGG CGTCCACTCG GAGAATCTCT TGTCTTACGA TATGCACCTT    1140
CACCACGANC GGGGCCCCAT GTNCNAGGAG CTCAATGCGT TTTTTCATAA CTGAGACTTC    1200
GCGCCGNCTC CCTNCTTTTT CTTTTGCCTT TGCCCGCCCC CCTGTCCCCA GCCCCCAGAG    1260
CGCAGGGACA CCCCCATNCT ACCCCGGCNC CGGCGGAGCG GGCCACCGGT CTGCCGCTCT    1320
CCTGGGGCAG CGCAGTCTGT TACNTGTGGT GG                                   1352
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 379 amino acids
(B) TYPE: amino acid

-continued

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Leu Thr Arg Leu Phe Ser Glu Pro Gly Leu Leu Ser Asp Val Pro
1 5 10 15

Lys Phe Ala Ser Trp Gly Asp Gly Glu Asp Asp Glu Pro Arg Ser Asp
20 25 30

Lys Gly Asp Ala Pro Pro Pro Pro Pro Pro Pro Ala Pro Gly Pro Gly Ala
35 40 45

Pro Gly Pro Ala Arg Ala Ala Lys Pro Val Pro Leu Arg Gly Glu Glu
50 55 60

Gly Thr Glu Ala Thr Leu Ala Glu Val Lys Glu Glu Gly Glu Leu Gly
65 70 75 80

Gly Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Gly Leu Asp Glu Ala
85 90 95

Glu Gly Glu Arg Pro Lys Lys Arg Gly Pro Lys Lys Arg Lys Met Thr
100 105 110

Lys Ala Arg Leu Glu Arg Ser Lys Leu Arg Arg Gln Lys Ala Asn Ala
115 120 125

Arg Glu Arg Asn Arg Met His Asp Leu Asn Ala Ala Leu Asp Asn Leu
130 135 140

Arg Lys Val Val Pro Cys Tyr Ser Lys Thr Gln Lys Leu Ser Lys Ile
145 150 155 160

Glu Thr Leu Arg Leu Ala Lys Asn Tyr Ile Trp Ala Leu Ser Glu Ile
165 170 175

Leu Arg Ser Gly Lys Arg Pro Asp Leu Val Ser Tyr Val Gln Thr Leu
180 185 190

Cys Lys Gly Leu Ser Gln Pro Thr Thr Asn Leu Val Ala Gly Cys Leu
195 200 205

Gln Leu Asn Ser Arg Asn Phe Leu Thr Glu Gln Gly Arg Asp Gly Ala
210 215 220

Xaa Arg Phe His Gly Ser Gly Gly Pro Phe Ala Met His Pro Tyr Pro
225 230 235 240

Tyr Pro Cys Ser Arg Gly Gly Arg Thr Val Pro Gly Ala Ala Ala Trp
245 250 255

Ala Ala Ala Gly Ala Arg Leu Arg Thr His Gly Tyr Cys Ala Ala Tyr
260 265 270

Glu Thr Leu Tyr Ala Ala Ala Gly Gly Gly Gly Ala Ser Pro Asp Tyr
275 280 285

Asn Ser Ser Glu Tyr Glu Gly Pro Leu Ser Pro Pro Leu Cys Leu Asn
290 295 300

Gly Asn Phe Ser Leu Lys Gln Asp Ser Ser Pro Asp His Glu Lys Ser
305 310 315 320

Tyr His Tyr Ser Met His Tyr Ser Gly Cys Pro Gly Ser Arg His Gly
325 330 335

His Gly Leu Val Phe Gly Ser Ser Ala Val Arg Gly Gly Val His Ser
340 345 350

Glu Asn Leu Leu Ser Tyr Asp Met His Leu His His Xaa Arg Gly Pro
355 360 365

Met Xaa Xaa Glu Leu Asn Ala Phe Phe His Asn
370 375

-continued

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 310 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: 20A1

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..310

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
CCCGGGCGTN CTGAGGTCCA GGGGCACAGG ACGACGAGCA GGAGAGGCGG CGGCGGCCGG      60

ACGCGNGTCC CTCCGAGGCG CTGCTGCACN CGCTGCGCAG GAGCGGCGCG TCAAGGCCAA     120

CGATCGCGAG CGCAACCGCA TGCACAACTT GAACGCGGCC CTGGACGCAC TGCGCAGCGT     180

GCTGCCCTCG TTCCCCGACG ACACCAAGCT CACCAAAATC GAGAGCCTGC GTTNCGCCTA     240

CAACTACATC TGGGCTCTGG CCGAGACACT GCGCTGGCGG ATNAAGGGCT GCCCGGAGGC     300

GGTGCCCGGG                                                            310
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 103 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Pro Gly Val Leu Arg Ser Arg Gly Thr Gly Arg Arg Ala Gly Glu Ala
1               5                  10                  15

Ala Ala Ala Gly Arg Xaa Ser Leu Arg Gly Ala Ala Ala Xaa Ala Ala
            20                  25                  30

Gln Glu Arg Arg Val Lys Ala Asn Asp Arg Glu Arg Asn Arg Met His
        35                  40                  45

Asn Leu Asn Ala Ala Leu Asp Ala Leu Arg Ser Val Leu Pro Ser Phe
    50                  55                  60

Pro Asp Asp Thr Lys Leu Thr Lys Ile Glu Ser Leu Arg Xaa Ala Tyr
65                  70                  75                  80

Asn Tyr Ile Trp Ala Leu Ala Glu Thr Leu Arg Trp Arg Xaa Lys Gly
                85                  90                  95

Cys Pro Glu Ala Val Pro Gly
            100
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1560 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

-continued

(i i) MOLECULE TYPE: cDNA

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE:
(B) CLONE: HC2A

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 57..1126

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
TTTTTCTGCT TTTCTTTCTG TTTGCCTCTC CCTTGTTGAA TGTAGGAAAT CGAAACATGA      60
CCAAATCGTA CAGCGAGAGT GGGCTGATGG GCGAGCCTCA GCCCCAAGGT CCTCCAAGCT     120
GGACAGACGA GTGTCTCAGT TCTCAGGACG AGGAGCACGA GGCAGACAAG AAGGAGGACG     180
ACCTCGAAGC CATGAACGCA GAGGAGGACT CACTGAGGAA CGGGGGAGAG GAGGAGGACG     240
AAGATGAGGA CCTGGAAGAG GAGGAAGAAG AGGAAGAGGA GGATGACGAT CAAAAGCCCA     300
AGAGACGCGG CCCCAAAAAG AAGAAGATGA CTAAGGCTCG CCTGGAGCGT TTTAAATTGA     360
GACGCATGAA GGCTAACGCC CGGGAGCGGA ACCGCATGCA CGGACTGAAC GCGGCGCTAG     420
ACAACCTGCG CAAGGTGGTG CCTTGCTATT CTAAGACGCA GAAGCTGTCC AAAATCGAGA     480
CTCTGCGCTT GGCCAAGAAC TACATCTGGG CTCTGTCGGA GATCCTGCGC TCAGGCAAAA     540
GCCCAGACCT GGTCTCCTTC GTTCAGACGC TTTGCAAGGG CTTATCCCAA CCCACCACCA     600
ACCTGGTTGC GGGCTGCCTG CAACTCAATC CTCGGACTTT TCTGCCTGAG CAGAACCAGG     660
ACATGCCCCC GCACCTGCCG ACGGCCAGCG CTTCCTTCCC TGTACACCCC TACTCCTACC     720
AGTCGCCTGG GCTGCCCAGT CCGNCTTACG GTACCATGGA CAGCTCCCAT GTCTTCCACG     780
TTAAGCCTCC GCCGCACGCC TACAGCGCAG CGCTGGAGCC CTTCTTTGAA AGCCCTCTGA     840
CTGATTGCAC CAGCCCTTCC TTTGATGGAC CCCTCAGCCC GCCGCTCAGC ATCAATGGCA     900
ACTTCTCTTT CAAACACGAA CCGTCCGCCG AGTTTGAGAA AAATTATGCC TTTACCATGC     960
ACTATCCTGC AGCGACACTG GCAGGGGCCC AAAGCCACGG ATCAATCTTC TCAGGCACCG    1020
CTGCCCCTCG CTGCGAGATC CCCATAGACA ATATTATGTC CTTCGATAGC CATTCACATC    1080
ATGAGCGAGT CATGAGTGCC CAGCTCAATG CCATATTTCA TGATTAGAGG CACGCCAGTT    1140
TCACCATTTC CGGGAAACGA ACCCACTGTG CTTACAGTGA CTGTCGTGTT TACAAAAGGC    1200
AGCCCTTTGG TACTACTGCT GCAAAGTGCA AATACTCCAA GCTTCAAGTG ATATATGTAT    1260
TTATTGTCAT TACTGCCTTT GGAAGAAACA GGGGATCAAA GTTCCTGTTC ACCTTATGTA    1320
TTATTTTCTA TAGACTCTTC TATTTTAAAA AATAAAAAAA TACAGTAAAG TTTAAAAAAT    1380
ACACCACGAA TTTGGTGTGG CTGTATTCAG ATCGTATTAA TTATCTGATC GGGATAACAA    1440
AATCACAAGC AATAATTAGG ATCTATGCAA TTTTTAAACT AGTAATGGGC CAATTAAAAT    1500
ATATATAAAT ATATATTTCA ACCAGCATTT TACTACTTGT TACCTCCCAT GCTGAATTAT    1560
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 356 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Met Thr Lys Ser Tyr Ser Glu Ser Gly Leu Met Gly Glu Pro Gln Pro
1 5 10 15

Gln Gly Pro Pro Ser Trp Thr Asp Glu Cys Leu Ser Ser Gln Asp Glu
20 25 30

Glu His Glu Ala Asp Lys Lys Glu Asp Asp Leu Glu Ala Met Asn Ala
35 40 45

Glu Glu Asp Ser Leu Arg Asn Gly Gly Glu Glu Glu Asp Glu Asp Glu
50 55 60

Asp Leu Glu Glu Glu Glu Glu Glu Glu Glu Glu Asp Asp Asp Gln Lys
65 70 75 80

Pro Lys Arg Arg Gly Pro Lys Lys Lys Lys Met Thr Lys Ala Arg Leu
85 90 95

Glu Arg Phe Lys Leu Arg Arg Met Lys Ala Asn Ala Arg Glu Arg Asn
100 105 110

Arg Met His Gly Leu Asn Ala Ala Leu Asp Asn Leu Arg Lys Val Val
115 120 125

Pro Cys Tyr Ser Lys Thr Gln Lys Leu Ser Lys Ile Glu Thr Leu Arg
130 135 140

Leu Ala Lys Asn Tyr Ile Trp Ala Leu Ser Glu Ile Leu Arg Ser Gly
145 150 155 160

Lys Ser Pro Asp Leu Val Ser Phe Val Gln Thr Leu Cys Lys Gly Leu
165 170 175

Ser Gln Pro Thr Thr Asn Leu Val Ala Gly Cys Leu Gln Leu Asn Pro
180 185 190

Arg Thr Phe Leu Pro Glu Gln Asn Gln Asp Met Pro Pro His Leu Pro
195 200 205

Thr Ala Ser Ala Ser Phe Pro Val His Pro Tyr Ser Tyr Gln Ser Pro
210 215 220

Gly Leu Pro Ser Pro Xaa Tyr Gly Thr Met Asp Ser Ser His Val Phe
225 230 235 240

His Val Lys Pro Pro Pro His Ala Tyr Ser Ala Ala Leu Glu Pro Phe
245 250 255

Phe Glu Ser Pro Leu Thr Asp Cys Thr Ser Pro Ser Phe Asp Gly Pro
260 265 270

Leu Ser Pro Pro Leu Ser Ile Asn Gly Asn Phe Ser Phe Lys His Glu
275 280 285

Pro Ser Ala Glu Phe Glu Lys Asn Tyr Ala Phe Thr Met His Tyr Pro
290 295 300

Ala Ala Thr Leu Ala Gly Ala Gln Ser His Gly Ser Ile Phe Ser Gly
305 310 315 320

Thr Ala Ala Pro Arg Cys Glu Ile Pro Ile Asp Asn Ile Met Ser Phe
325 330 335

Asp Ser His Ser His His Glu Arg Val Met Ser Ala Gln Leu Asn Ala
340 345 350

Ile Phe His Asp
355

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1462 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

-continued

(i i) MOLECULE TYPE: cDNA

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Mus musculus

(v i i) IMMEDIATE SOURCE:
(B) CLONE: 1.1.1

(i x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 231..1101

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
GAATTCAAGC TAGAGGCTGG TACCCCGCCT GGTAGAGATG CCACACTCGC TCCGCGGCTC      60
GCATGGCGCT CTGAAGACGC CGGCGCCCGC CGCCTTGAGG AACCGCTGCC CCCGCTCCCT     120
GAAGATGGGG GAACAATGAA ATAAGCGAGA AGATTCCTCT TCTCCCCCCT CTCTCTCTTG     180
CCCCCTCCCC CCTCCCCTCC CCTCTCCCCT TGACTCCTCT CTGAGGCACC ATGCTGACCC     240
GCCTGTTCAG CGAGCCCGGC CTCCTCTCGG ACGTGCCCAA GTTCGCCAGC TGGGGCGACG     300
GCGACGACGA CGAGCCGAGG AGCGACAAGG GCGACGCGCC GCCGCAGCCT TCTCCTGCTC     360
CCGGGTCGGG GGCTCCAGGA CCCGCCCGGG CCGCCAAGCC AGTGTCTCTT CGTGGAGGAG     420
AAGAGATCCC TGAACCCACG TTGGCTGAGG TCAAGGAGGA AGGAGAGCTG GGCGGCGAGG     480
AGGAGGAGGA AGAGGAGGAG GAGGAAGGAC TGGACGAGGC GGAAGGCGAG CGGCCCAAGA     540
AGCGCGGGCC GAAGAAACGC AAGATGACCA AGGCGCGTCT GGAGCGCTCC AAGCTGCGGC     600
GACAGAAGGC CAATGCGCGC GAGCGCAACC GCATGCACGA CCTGAACGCG GCTCTGGACA     660
ACCTGCGCAA GGTGGTCCCC TGCTACTCCA AGACCCAGAA GCTGTCCAAG ATCGAGACCC     720
TGCGCCTGGC CAAGAACTAC ATCTGGGCTC TCTCGGAGAT CTTGCGCTCC GGGAAGCGGC     780
CGGATCTGGT GTCCTACGTG CAGACTCTGT GCAAGGGGCT GTCACAGCCC ACCACGAATC     840
TGGTGGCCGG CTGCCTGCAG TTAAACTCTC GTAACTTCCT CACGGAGCAG GGCGCGGACG     900
GCGGCCGCTT TCACGGCTCG GGTGGCCCGT TCGCCATGCA TCCGTACCCA TACCCGTGCT     960
CCCGCCTGGC AGGCCACAGT GTCAGGCGGC TGGCGGCCTG GGCGGAGGNC GGCGCACGCC    1020
TGCGGACCCA CGGCTACTGC GCCGCCTACG AGACGCTGTA CGCGGCGGCC GGTGGCGGCG    1080
GCGCTAGCCC GGACTACAAC AGCTCCGAGT ACGAGGGTCC ACTCAGTCCC CCGCTCTGTC    1140
TCAACGGCAA CTTCTCGCTC AAGCAGGACT CGTCCCCCGA TCACGAGAAG AGCTACCACT    1200
ACTCTATGCA CTACTCGCGC TGCCCNGGCT CACGCCACGG NCACGGGCTG GTCTTCGGCT    1260
CGTCGGCCGT GCGCGGGGGC GTCCACTCCG AGAATCTCTT GTCTTACGAT ATGCACCTTC    1320
ACCACGATCG GGGCCCCATG TACGAGGAGC TCAACGCATT TTTCCATAAC TGAGACCTCN    1380
CGCCGACCCC TTCTTTTTCT TTGCCTTNNT CCGGCCCCTT AGCCCCANCC CCAANANCTC    1440
AGGNNTCCCA CCGATCTCCA GG                                             1462
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 380 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(v i) ORIGINAL SOURCE:
(A) ORGANISM: Mus musculus

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met Leu Thr Arg Leu Phe Ser Glu Pro Gly Leu Leu Ser Asp Val Pro

-continued

1				5						10					15		
Lys	Phe	Ala	Ser	Trp	Gly	Asp	Gly	Asp	Asp	Asp	Glu	Pro	Arg	Ser	Asp		
			20					25					30				
Lys	Gly	Asp	Ala	Pro	Pro	Gln	Pro	Ser	Pro	Ala	Pro	Gly	Ser	Gly	Ala		
		35					40					45					
Pro	Gly	Pro	Ala	Arg	Ala	Ala	Lys	Pro	Val	Ser	Leu	Arg	Gly	Gly	Glu		
	50					55					60						
Glu	Ile	Pro	Glu	Pro	Thr	Leu	Ala	Glu	Val	Lys	Glu	Glu	Gly	Glu	Leu		
65					70					75					80		
Gly	Gly	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Gly	Leu	Asp	Glu		
				85					90					95			
Ala	Glu	Gly	Glu	Arg	Pro	Lys	Lys	Arg	Gly	Pro	Lys	Lys	Arg	Lys	Met		
			100					105					110				
Thr	Lys	Ala	Arg	Leu	Glu	Arg	Ser	Lys	Leu	Arg	Arg	Gln	Lys	Ala	Asn		
		115					120					125					
Ala	Arg	Glu	Arg	Asn	Arg	Met	His	Asp	Leu	Asn	Ala	Ala	Leu	Asp	Asn		
	130					135					140						
Leu	Arg	Lys	Val	Val	Pro	Cys	Tyr	Ser	Lys	Thr	Gln	Lys	Leu	Ser	Lys		
145					150					155					160		
Ile	Glu	Thr	Leu	Arg	Leu	Ala	Lys	Asn	Tyr	Ile	Trp	Ala	Leu	Ser	Glu		
				165					170					175			
Ile	Leu	Arg	Ser	Gly	Lys	Arg	Pro	Asp	Leu	Val	Ser	Tyr	Val	Gln	Thr		
			180					185					190				
Leu	Cys	Lys	Gly	Leu	Ser	Gln	Pro	Thr	Thr	Asn	Leu	Val	Ala	Gly	Cys		
		195					200					205					
Leu	Gln	Leu	Asn	Ser	Arg	Asn	Phe	Leu	Thr	Glu	Gln	Gly	Ala	Asp	Gly		
	210					215					220						
Gly	Arg	Phe	His	Gly	Ser	Gly	Gly	Pro	Phe	Ala	Met	His	Pro	Tyr	Pro		
225					230					235					240		
Tyr	Pro	Cys	Ser	Arg	Leu	Ala	Gly	His	Ser	Val	Arg	Arg	Leu	Ala	Ala		
				245					250					255			
Trp	Ala	Glu	Xaa	Gly	Ala	Arg	Leu	Arg	Thr	His	Gly	Tyr	Cys	Ala	Ala		
			260					265					270				
Tyr	Glu	Thr	Leu	Tyr	Ala	Ala	Ala	Gly	Gly	Gly	Gly	Ala	Ser	Pro	Asp		
		275					280					285					
Tyr	Asn	Ser	Ser	Glu	Tyr	Glu	Gly	Pro	Leu	Ser	Pro	Pro	Leu	Cys	Leu		
	290					295					300						
Asn	Gly	Asn	Phe	Ser	Leu	Lys	Gln	Asp	Ser	Ser	Pro	Asp	His	Glu	Lys		
305					310					315					320		
Ser	Tyr	His	Tyr	Ser	Met	His	Tyr	Ser	Arg	Cys	Pro	Gly	Ser	Arg	His		
				325					330					335			
Gly	His	Gly	Leu	Val	Phe	Gly	Ser	Ser	Ala	Val	Arg	Gly	Gly	Val	His		
			340					345					350				
Ser	Glu	Asn	Leu	Leu	Ser	Tyr	Asp	Met	His	Leu	His	His	Asp	Arg	Gly		
		355					360					365					
Pro	Met	Tyr	Glu	Glu	Leu	Asn	Ala	Phe	Phe	His	Asn						
	370					375					380						

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

-continued

(i i) MOLECULE TYPE: cDNA

(v i i) IMMEDIATE SOURCE:
(B) CLONE: JL34

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTCAGCATCA GCAACTCGGC 20

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(v i i) IMMEDIATE SOURCE:
(B) CLONE: JL36

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TCGGATCCCG TTCTAGGCGC GCCTTGGTC 29

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(v i i) IMMEDIATE SOURCE:
(B) CLONE: JL40

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTTTTCCCAG TCACGACGTT G 21

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. An isolated nucleic acid molecule which hybridizes under stringent conditions with a nucleic acid molecule selected from among SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, and complements thereof.

2. A vector comprising in serial array a promoter, and the nucleic acid molecule of claim 1.

3. A cell in culture transformed by the nucleic acid molecule of claim 1.

4. A method for inducing differentiation of a non-neuronal cell in culture into a neuron, comprising introducing a nucleic acid molecule of claim 1 into the non-neuronal cell.

5. An isolated nucleic acid molecule, wherein the nucleic acid molecule encodes a polypeptide having an amino acid sequence selected from among the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, and SEQ ID NO:17.

6. A vector comprising in serial array a promoter, and the nucleic acid molecule of claim 5.

7. A cell in culture transformed by the nucleic acid molecule of claim 5.

8. A method for inducing differentiation of a non-neuronal cell in culture into a neuronal cell in culture into a neuronal cell, comprising introducing the nucleic acid molecule of claim 5, into the non-neuronal cell.

* * * * *